

Novel Aspects of RNA Synthesis and Regulation:

**Non-Canonical RNA Capping by DNA-Dependent
RNA Polymerases**

&

**Screen for Bacteriophage Regulators of
Mycobacterium smegmatis Transcription**

Christina Julius

A dissertation submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Centre for Bacterial Cell Biology

Institute for Cell and Molecular Biosciences

Medical School

Newcastle University

October 2019



Centre For Bacterial Cell Biology

I. Abstract

In the process of gene expression, all living cells synthesise RNA. RNA is polymerised from nucleotide monomers in dependence of a complementary sequence on a DNA template, a gene. This is done by DNA-dependent RNA Polymerase (RNAP) and is called transcription.

Previously, it was believed that only eukaryotic RNA can be 5'-"capped". The 5'-cap in eukaryotes is attached post-transcriptionally and serves several functions in RNA metabolism and in translation (process of protein synthesis from an RNA template). Recently, a cap was found on bacterial RNA. The bacterial cap is not a modified canonical nucleotide (like the methylated guanosine cap of eukaryotes), but consists of small metabolites or cell wall precursors which are covalently bound to 5'-RNA. This thesis investigates how bacterial RNAP incorporates the non-canonical initiating nucleotide (NCIN)-cap and discusses structural determinants of NCIN-capping of both RNAP enzyme and template DNA. RNAP accepts the cofactor and cell wall precursor as initiating substrate due to the free 3'-hydroxyl group of the nucleotide moiety of the molecule. Functional moieties such as nicotinamide ribose or riboflavin on the 5'-end of the nucleotide ribose are bound as substrate by wildtype RNAP catalytic centre and RNA exit channel, and do not interact with upstream promoter regions. However, larger nucleotide-containing molecules such as uridine diphosphate N-acetylmuramic acid pentapeptide are inefficiently incorporated due to sigma subunit of RNAP.

We further show that primase, RNA-polymerase of replication, incorporates NCIN into primer RNA and explore potential downstream effects on replication. DNA Polymerase I which replaces primer RNA with DNA, is differentially affected by 5'-cofactors on primer RNA.

In the bigger picture, NCIN-caps as extensions of RNA primary structure might have a variety of physiological consequences, which are as of yet uncertain, but extensively discussed in this thesis. We argue that RNA stability, translation, localisation, and activity of regulatory RNA might be affected.

RNA synthesis regulation in bacteria makes a fantastic target for drug development, as the cellular machineries of prokaryotes and eukaryotes differ. Also bacteria-infecting viruses, or bacteriophages, are often reported to interfere with host transcription in order to monopolise cellular resources in favour of their own gene expression and replication. We propose that studying the mechanisms by which bacteriophages hijack host gene expression could lead to the discovery or development of new antimicrobials. In order to investigate novel aspects of transcription regulation, we outline a methodology to study how bacteriophages regulate host transcription in mycobacteria. This bacterial family includes today's most deadly bacterial pathogen, *Mycobacterium tuberculosis*. At the example of mycobacteriophage D29 gene product Gp53, we show that the protocol can be used to find new inhibitors of bacterial transcription.

II. Acknowledgements

I want to thank first and foremost my supervisor Dr Yulia Yuzenkova who has taught me many things, supported me whenever needed, and allowed me to develop independence as a researcher. She has further given me the amazing opportunity to join her studies on NCIN-capping. I feel very lucky to have taken part in discovery of something unexpected and to have gotten the invaluable chance to publish our results. I also want to thank other (former) members of the Zenkin/Yuzenkova group for professional and moral support and many positive memories, most of all Dr Amber Riaz-Bradley for much help and purification of NudC.

Further, I want to thank Dr Kevin Waldron from Newcastle University for performing ICP-MS and Dr Matthew Peake from Newcastle University for strain JW4012-AM (DnaB). With gratitude we received UDP-Glc and UDP-GlcNAc from Prof Waldemar Vollmer.

I thank all technical and administrative staff of ICaMB and CBCB for many working hours.

Outside of Newcastle University, I would like to acknowledge our collaborators Prof Graham Hatful from Pittsburgh University, who provided *M. smegmatis* strains and mycobacteriophages, and Deborah Jacobs-Sera from the Hatful Lab for advice on *M. smegmatis* growth and infection. I want to thank Prod Dipankar Chatterji for *M. smegmatis* wt and SM07. I further thank Dr Andrey Revyakin for proteins POLMRT, TFAM and TFB2M.

I also want to thank my parents Monika and Norbert Julius for always believing in me and neglecting to teach me that there are things that girls cannot do. I am grateful for all the sacrifice that parenthood entails on so many levels. I want to thank my sister, Miriam Julius for encouragement and friendship and my brother Lucas Julius for fantastic entertainment.

III. Abbreviations

Abbreviation	Meaning	Abbreviation	Meaning
AMP	adenosine monophosphate	NAD(H)(P)	nicotinamide adenine dinucleotide (oxidised or reduced form)(-phosphate)
ATP	adenosine triphosphate	NCIN	non-canonical initiating nucleotide
%	percent	NCIS	non-canonical initiating substrate
μ	micro	Ni	nickle
A	Ampere	NMN	nicotinamide mononucleotide
bp	base pair(s)	NMP	nucleoside monophosphate
C	carbon	nt	nucleotide(s)
C	Celsius	N-terminal/terminus	amino-terminal/terminus
Ci	Curie	NTP	nucleoside triphosphate
Cl	chloride	NUDIX	nucleotide diphosphate attached to a moiety X
Co	cobalt	O	oxygen
C-terminal/terminus	carboxyl-terminal/terminus	OD600	optical density at a wavelength of 600 nm
CTP	cytidine triphosphate	ORF	open reading frame
Cu	copper	p	pico
Da	Dalton	P1/2	promoter 1/2
DNA	deoxyribonucleic acid	PAGE	polyacrylamide Gel Electrophoresis
DP-CoA	dephospho-Coenzyme A	PCR	polymerase chain reaction
DR	drug resistant	Ppi	pyrophosphate
DS	drug-sensitive	ppm	parts per million
dsDNA	double stranded DNA	Pt	platin
DTT	dithiothreitol	PVDF	polyvinylidene difluoride
Ec	<i>E. coli</i>	rcf	relative centrifugal force
EDTA	ethylenediaminetetraacetic acid	Rif-pocket	Rifampicin-binding pocket
FAD	flavin adenine dinucleotide	RNA	ribonucleic acid
Fe	iron	RNAP	DNA-dependent RNA polymerase
FMN	flavin mononucleotide	RPc	closed promoter complex

FPLC	fast protein liquid chromatography	RPinit	initially transcribing RNAP-promoter complex
g	gram	rpm	revolutions per minute
Gp	gene product	RPo	open promoter complex
GTP	guanosine triphosphate	rRNA	ribosomal RNA
H	hydrogen	S	sulfur (unless mention in conjunction to RNA name or size)
h	hour	s	second
In	iodine	scaRNA	small Cajal body-specific RNA
IPTG	isopropyl β -D-1-thiogalactopyranoside	SDS	sodium dodecylsulphate
k	kilo	snoRNA	small nucleolar RNA
kb	kilobase	sRNA	small regulatory RNA
l	litre	ssDNA	single stranded DNA
LB	Luria Bertani	SVP	snake venom phosphodiesterase
LC	liquid chromatography	t	time
M	Molar concentration	TB	tuberculosis
m	milli	TL	trigger loop
MDR	multi-drug resistant	tRNA	transfer RNA
Mg	magnesium	TSS/ +1	transcription start site
Mn	manganese	UDP-Glc	uridine diphosphate glucose
mRNA	messenger RNA	UDP-GlcNAc	uridine diphosphate N-acetylglucosamine
Ms	<i>M. smegmatis</i>	UDP-MurNAc-AA5	uridine diphosphate N-acetylmuramic acid pentapeptide
Mtb	<i>M. tuberculosis</i>	UTP	uracil triphosphate
MWCO	molecular weight cut off	V	Volts
N	nitrogen	W	Watt
n	nano (unless in conjunction with nucleobasis)	WHO	World Health Organisation
		wt	wildtype
		XDR	extensively drug resistant
		Zn	zinc

IV. Table of Contents

I.	Abstract	I
II.	Acknowledgements	III
III.	Abbreviations	V
IV.	Table of contents	VII
V.	Figures and tables	X
1.	General Introduction	1
1.1	Reasoning of this study	1
1.2	The basis of RNA function is its complex and versatile structure	2
1.3	RNA degradation regulates gene expression	5
1.4	Many abundant small-molecule metabolites contain a nucleotide-moiety ..	7
1.5	Transcription is a central cellular process	8
1.5.1	RNA polymerase structure of bacteria and eukaryotes is similar	8
1.5.2	RNA polymerase catalysis is a two metal ion mechanism	11
1.5.3	Transcription is a series of distinct steps from initiation over elongation to termination	15
2.	Materials and Methods	19
2.1.1	Supplier lists	19
2.1.2	Recipes	22
2.1.3	Strains, primer and templates	25
2.1.4	Sequences (vectors, templates and primers)	26
2.2	Microbiological methods	32
2.2.1	Bacterial growth and storage	32
2.2.2	Mycobacteriophage growth, infection and storage	33
2.3	Molecular cloning	35
2.4	Protein purification	39
2.5	NudC metal analysis	44
2.6	<i>In vitro</i> RNA synthesis	44
2.6.1	Synthesis of transcription or priming templates	44
2.6.2	<i>In vitro</i> transcription with <i>E. coli</i> RNA polymerase from <i>E. coli</i> or T7 phage promoters	45
2.6.3	NCIN-decapping with NudC	45
2.6.4	NCIN-capping by POLMRT and decapping by NudC	46
2.6.5	<i>In vitro</i> transcription with <i>M. smegmatis</i> RNA polymerase from <i>M. smegmatis</i> or L5 phage promoters	47
2.6.6	<i>In vitro</i> primer synthesis by DnaG	47
2.6.7	NCIS versus ppGpp competition assay	47
2.6.8	Primer decapping by NudC	48
2.6.9	Primer cleavage by RNase H, SVP, or RppH	48
2.6.10	Primer removal by Pol I	48
2.6.11	Primer initiation by P49	48
2.6.12	Denaturing bis:acrylamide gel electrophoresis	49

3	Chapter 1: Bacterial and mitochondrial RNA Polymerases perform non-canonical RNA capping	51
3.2	Introduction	51
3.2.1	Chronology of the Discovery	51
3.2.2	Mitochondrial transcription is dissimilar to bacterial and eukaryotic transcription	57
3.3	Aims	60
3.4	Results	60
3.4.1	Bacterial RNA polymerase incorporates adenine containing cofactors for transcription on +1A promoters	60
3.4.2	Uridine-containing cell wall precursors initiate transcription on +1U promoters	63
3.4.3	Promoter -1 position does not affect NADylation efficiency	64
3.4.4	σ factor might the limit size of the NCIN-cap	65
3.4.5	Rif-pocket residues might interact with NCIN-cap during initiation	67
3.4.6	NUDIX hydrolase NudC acts as decapping enzyme <i>in vitro</i> and can be used as a tool to verify NCIN-cap	69
3.4.7	NudC purification shows that NudC replaces Zn^{2+} with Fe^{2+}	71
3.4.8	Mitochondrial RNA polymerase incorporates nucleotide-analogues as initiating nucleotides	71
3.5	Discussion	72
3.5.1	Metabolites and cell wall precursors on 5'-RNA were identified <i>in vivo</i> and shown to be incorporated by RNAP <i>in vitro</i>	72
3.5.2	Promoter-mediated regulation of NCIN-capping is controversial	73
3.5.3	Protein structural determinants of NCIN-capping were found in σ and β subunits of EcRNAP	75
3.5.4	NCIN-capping stimulates promoter escape	76
3.5.5	Mitochondrial single-subunit RNAP readily incorporates adenosine-containing metabolites as 5'-caps	77
3.5.6	NCIN-caps can be removed by NUDIX hydrolases	78
3.5.7	Metabolite-caps might be incorporated post-transcriptionally	81
3.5.8	NCIN-capping might have various physiological implications	82
3.6	Conclusion and Outlook	90
4	Chapter II: Primase utilises cellular cofactors to initiate primer synthesis ..	91
4.2	Introduction	91
4.2.1	DnaG is an RNA polymerase that looks like a topoisomerase	92
4.2.2	RNA primer synthesis by DnaG continually initiates DNA synthesis	94
4.2.3	Replication is regulated in stringent response via regulation of DnaG	97
4.2.4	Eukaryotic primer synthesis and removal differ from the prokaryotic system	98
4.3	Aims	102
4.4	Results	103
4.4.1	DnaG initiates primer synthesis with metabolic cofactors NAD(H), FAD and DP-CoA at varying efficiency	103
4.4.2	DnaB improves specificity of primer initiation with canonical and non-canonical substrates	103
4.4.3	NudC can hydrolyse 5'-NADH on primer	103
4.4.4	Cofactor-modified 5'-primer RNA does not interact with DnaG recognition sequence -1 position on DNA	106

4.4.5	DnaG basic ridge residues are involved in binding of non-canonical substrate	107
4.4.6	Stringent response alarmone ppGpp inhibits NCIS-primer initiation	108
4.4.7	5'-cofactors differentially affect primer removal by DNA polymerase I ...	108
4.4.8	5'-cofactors do not affect primer removal by RNase H	111
4.4.9	The human primase catalytic subunit P49 can initiate primer synthesis with guanosine-analogues	112
4.5	Discussion	113
4.5.1	Bacterial primase DnaG readily initiates primer synthesis using nucleotide-containing cofactors of metabolism	113
4.5.2	NCIS incorporation by DnaG might have downstream effects on replication	114
4.5.3	Starvation response results in inhibition of NCIS-primer synthesis	117
4.5.4	Eukaryotic primase might initiate primer synthesis with non-canonical substrates	117
4.6	Conclusion and outlook	120
5.	Chapter III: Screen for Bacteriophage Regulators of <i>Mycobacterium smegmatis</i> Transcription	121
5.1	Introduction	121
5.1.1	Mycobacteriophages are diverse and resourceful	121
5.1.2	Tuberculosis treatment requires research and innovation	123
5.1.3	Mycobacteriophage therapy could aid TB treatment	126
5.1.4	Enzybiotics are protein antibacterials of phage origin	127
5.1.5	Host transcription regulation is a common phage strategy	129
5.2	Aims	134
5.3	Results and Discussion	134
5.3.1	Workflow	134
5.3.2	Optimisation of <i>M. smegmatis</i> growth conditions	135
5.3.3	Phage titration methods	136
5.3.4	Determination of an infection timeline	137
5.3.5	Cell harvesting	139
5.3.6	Phage protein pull-down via His-tagged mycobacterial RNA Polymerase .	140
5.3.7	Mass Spectrometry analysis of pull-down	142
5.3.8	Cloning of phage proteins	143
5.3.9	Purification of phage proteins	143
5.3.10	His-tagged Gp53 supports transcription of long products	144
5.3.11	Gp53-mediated transcription inhibition relies on free N-terminal domain .	145
5.3.12	Gp53 inhibits <i>M. smegmatis</i> RNAP, but not <i>E. coli</i> RNAP	147
5.3.13	Gp53 interacts with RNAP rather than DNA	148
5.3.14	Further improvement of the protocol	149
5.4	Conclusion	150
5.5	Outlook/ Perspectives	151
6.	References	153
7.	Bibliography	187

V. Figures and Tables

Figure 1.1 The bacterial degradosome	6
Figure 1.2 Molecular structures of ADP-containing molecules	9
Figure 1.3 RNAP subunits and domains	11
Figure 1.4 RNAP catalysis	14
Figure 1.5 Bacterial transcription initiation and transition into elongation	16
Table 2.1 Supplier of chemicals, consumables, kits and commercial media	19
Table 2.2 Supplier of enzymes	21
Table 2.3 Recipes of homemade media and buffer used in this study	22
Table 2.4 Strains of bacteria used in this study	25
Table 2.5 Primer (and purpose) used in these studies	26
Table 2.6 Templates for transcription and primer synthesis used in these studies	29
Table 2.7 Mycobacteriophages provided by Prof Graham Hatfull	33
Figure 2.1 D29 plaque lace pattern	34
Table 2.8 PCR recipes and PCR machine settings for Q5, Phusion and Taq polymerases	36
Figure 2.3 pET28a-p49 plasmid as constructed by Twist Bioscience	38
Table 3.1 Chronology of discoveries on NCIN-capping and decapping	54
Figure 3.1 Mitochondrial transcription initiation	59
Table 3.2 Comparison of cellular concentration and K_M of metabolites used as initiating substrates for transcription	61
Figure 3.2 Promoter-dependent transcription initiation by EcRNAP using different adenine-containing metabolites	62
Figure 3.3 Metabolite-capping on promoter <i>rnaI</i>	62
Table 3.3 Comparison of cellular concentration and K_M of UTP and cell wall precursors used as initiating substrates for transcription	63
Figure 3.4 Transcription initiation by uridine-containing molecules on promoter <i>acnA</i> .	64
Figure 3.5 Effect of template position -1 on incorporation of canonical substrate ATP vs non-canonical substrate NAD^+	65
Figure 3.6. Role of σ in NCIN-capping	66
Figure 3.7 Initiation with NAD^+ promotes promoter escape	68
Figure 3.8 Rif-pocket residues are involved in NCIN-capping	68
Figure 3.9. NudC decaps FAD-capped RNA at lower efficiency than NAD^+ -RNA	70
Figure 3.10 Accessibility of RNAP-bound NAD^+ -RNA products to NudC	70
Figure 3.11 NCIN-capping by POLMRT	72
Figure 3.12 Crystal structure of NudC in complex with NAD^+	79
Figure 3.13 Schematic of NCIN-capping and decapping	81
Figure 3.14 Theory of a post-transcriptionally active metabolite-capping enzyme	82
Table 3.4 Functional destinations of RNAs which have been identified as frequently NAD^+ -capped	84
Table 3.5 sRNAs which were found to be highly NAD ylated in <i>E.coli</i>	86
Figure 4.1. DnaG primase from <i>M. tuberculosis</i> , <i>S. aureus</i> and <i>E.coli</i> structures and interaction with cofactors, substrates and template	95
Figure 4.2 The bacterial replisome and trombone model	97
Figure 4.3 Stringent response and DnaG binding of alarmones	98
Figure 4.4. Eukaryotic and archaeal primase	101
Figure 4.5 DnaG utilises adenine-containing cofactors of metabolism to initiate primer synthesis on ssDNA	104

Figure 4.6 DnaB stimulates DnaG activity	105
Figure 4.7. NudC can be used to verify 5'-NADH-primer	105
Figure 4.8. -1 template base affects initiation efficiency with canonical and non-canonical substrates, but does not affect relative efficiency of cofactor incorporation. ...	106
Figure 4.9. DnaG residues of basic ridge are likely involved in NCIS binding	107
Figure 4.10 Inhibition of NCIN-primer synthesis by ppGpp	109
Figure 4.11 Primer removal by Pol I is affected by 5'-cofactors	110
Figure 4.12. Cleavage of primer by RNase H and decapping enzymes	112
Figure 4.13. Incorporation of guanosine-containing small molecules by eukaryotic primase catalytic subunit P49	112
Figure 5.1 Phage lifestyles	124
Figure 5.2 Gene expression by a hypothetical lytic mycobacteriophage	131
Table 5.1 Inhibitors of bacterial transcription of phage origin (phage effector) vs antibiotics or chemicals that are used in clinic or research	132
Figure 5.3. Schematic of infection workflow	134
Figure 5.4 Growth of <i>M. smegmatis</i> in liquid culture at different temperatures	136
Figure 5.5 Use of tyloxapol or xylose to prevent <i>M. smegmatis</i> cell aggregation	137
Figure 5.6 Spot test and plate assay for titration	138
Figure 5.7 Correlation of <i>M. smegmatis</i> growth and D29 titre in infected cultures	139
Table 5.2 Yield of purified His-tagged RNAP for different cell lysis methods	140
Figure 5.8 SDS-PAGE of RNAP pull-down from infected and uninfected cells	141
Table 5.3 Mycobacteriophage Gps found in samples of infected RNAP purifications ...	142
Figure 5.9 Purification of His-Gp53	144
Figure 5.10 Transcription activation by first Gp53 purification	146
Figure 5.11 Gp53 inhibits transcription on two promoters	147
Figure 5.12 Transcription inhibition of Gp53 is species-specific	148
Figure 5.13. Gp53 inhibits transcription via protein-binding	149
Figure 5.14 Structure of mycobacterial transcription factor RbpA in complex with MsRNAP holoenzyme and DNA	152

1. General Introduction

1.1 Reasoning of this study

The central dogma of life on earth is based on three main biomolecules: DNA, RNA, and protein. DNA is the blueprint of life bearing the genetic code. RNA is the messenger molecule or intermediate, presenting the genetic code in an accessible version. Protein is the realisation of the genes in form of a structural or catalytic molecule. Transcription is an essential cellular process that is executed by all living organisms as well as DNA viruses as part of gene expression. DNA is transcribed into an RNA copy which either has a function itself, or is further translated into protein. The central enzyme of transcription is DNA-dependent RNA Polymerase (RNAP), which matches ribonucleotides to complementary bases on the DNA template strand and polymerises them to form an RNA chain, the transcript. The process of transcription as well as the transcript's fate can be regulated in a variety of ways. An organism needs to regulate its own transcription in order to keep up with momentary requirements of a cell (e.g. production of enzymes to break down a specific nutrient the cell encounters), or interact with the environment (e.g. heatshock or infection response). In return, the organism can act upon other organisms in its environment, to modulate their transcription (e.g. host shut-off in infection). All chapters of my thesis were born from the studies of regulation of gene expression. However, they are vastly different in their focus. We initially intended to study how mycobacteriophages manage to hijack host transcription. This is an interesting study objective as it not only provides insight into viral and bacterial gene expression and potentially novel mechanisms of regulation, but this insight might also have important applications, first and foremost in development of new antibacterials against tuberculosis, which is a globally pressing issue. As these studies were beginning, discoveries were reported which showed a fascinating potential to research previously unnoticed mechanisms of regulation of gene expression via RNA synthesis and RNA primary structure. The phenomenon of non-canonical RNA capping might constitute an ancient mechanism as it has been found in eubacteria as well as eukaryotes. While the function of non-canonical capping is at the present time uncertain, part of the mechanism as well as hypothetical functions are discussed in chapter I. Chapter II is a spin-off project that is concerned with RNA synthesis in replication (primer synthesis), in contrast to transcription. Primase, despite being entirely unrelated to transcriptases, integrates non-canonical nucleotides into primer RNA, with potential downstream effects on replication. Chapter III describes development of a methodology for our lab that will enable us to further explore mycobacteriophage regulation

of host transcription in the future. After establishing this protocol it has been used by two master students in pursuit of phage regulators, and one phage inhibitor of transcription, Gp53 of mycobacteriophage D29, has been initially characterised.

1.2 The basis of RNA function is its complex and versatile structure

RNA has many functions in the cell, most prominently ribosomal RNA (rRNA), transfer RNA (tRNA), and messenger RNA (mRNA).

The majority of cellular RNA (in bacteria 83 % of total RNA dry mass, in mammalian cells 80%) is rRNA (Lodish *et al.*, 2004; Feijó Delgado *et al.*, 2013). The ribosome is a ribozyme, supported by a protein scaffold. It translates all cellular protein from mRNA. The ribosome consists of two subunits. In bacteria the 70S ribosome consists of 30S and 50S subunits which contain 16S and 23S rRNA, respectively, in association to ribosomal protein (Brimacombe & Stiege, 1985, Brodersen, Nissen, 2005). rRNA operons in *E.coli* are transcribed from two operons behind promoters *rrn* P1 and *rrn* P2 (Paul, Ross, *et al.*, 2004). Bacteria can have up to 15 different copies of the rRNA genes, with sequence variation up to 1 % (Leppek and Barna, 2019). In *E. coli*, 16S, 23S and 5S rRNA precursors are cleaved from an initial 30S polycistronic transcript by combined actions of RNase III, RNase E, RNase G and RNase T (Deutscher, 2009). The protein portion of the ribosome provides a scaffold for ribozyme structure, and has been proposed to be involved in mRNA and tRNA binding and interaction with translation factors (Brodersen, Nissen, 2005).

tRNA constitutes 15% of cellular RNA (Lodish *et al.*, 2004). It has a clover leaf structure with several conserved sequences and modified bases. An amino acid is enzymatically attached to the tRNA 3'-OH to give aminoacylated or activated tRNA. An anticodon loop in the tRNA body base pairs with the codon on mRNA. Thus, identity of amino acid and cognate anti-codon are vital to realise the genetic code. The ribosome polymerises outgoing and incoming amino acids from tRNA, releasing naked tRNA (Yarus, Caporaso and Knight, 2005).

mRNA is a “copy” of the genetic code which is translated into a peptide sequence. The 5'-end of mRNA certainly exerts the biggest effect on translation by regulating initiation via ribosome binding. But besides that, mRNA structure plays a role in protein folding and translation fidelity. The more stable secondary structure regions on mRNA, the slower translation will progress. This allows for correct folding of compacter protein regions (Faure

et al., 2016). Also, high structuring in mRNA can increase transcript stability to allow for more frequent translation. Faure *et al.* also demonstrated that “fast” (less structured) regions on mRNA typically encode α -helices, while “slow” (densely structured) regions on mRNA lead to formation of β -strands and loops on protein.

The basis of RNA versatility is RNA folding. RNA differs from DNA only by uracil/thymine base and by presence of 2'-OH in pentose (Westhof and Auffinger, 2000). 2'-OH allows for an increased flexibility of the RNA backbone as compared to DNA. Thus, besides base pairing to a complementary second nucleic acid, RNA folds in on itself. Various non-covalent interactions create secondary and tertiary structures (Batey, Rambo and Doudna, 1999).

Secondary structure is dependent on primary base sequence. Base pairing with interrupted ssRNA stretches can form to hairpins, loops, helical duplexes, internal loops, bulges, and junctions (Batey, Rambo and Doudna, 1999). Tertiary structure is assumed after processing via splicing (removing parts of the transcript) or methylation of nucleobases or 2'-OH on ribose (Batey, Rambo and Doudna, 1999), and includes longer-range interactions (tertiary interactions) between more distal regions on the molecule.

RNA structures include but are not limited to coaxial stacking, 2'-OH-mediated interactions, triple helices, tetraloop motifs, ribose zipper, metal-core motifs, kissing loops, and pseudoknots. Coaxial stacking is an alignment of a stack of helices or hairpin loops along a common axis in space, which improves stability. 2'-OH-mediated interactions are achieved when hydrogen bonds are made between 2'-OH, 3'-O, and/or phosphate oxygens to a N or O of a base. For example, a kink-turn motif arises from interactions of 2'-OH of neighbouring helices, or backbone to backbone helical interactions. Triple helices are composed of a third strand fitting into the major groove of double helix. Tetraloop motifs are free four-nucleotide sections within an otherwise base paired strand. Ribose zipper are interactions of a bulge between two tetraloops, backbone to backbone, with transient interactions between 2'-OH and N3 of a purine base and 2'-OH and O2 of a pyrimidine base creating a “zipper” within minor grooves of two helices. Metal-core motifs rely on tight packing of backbones to create electrostatic interactions which are released by binding of divalent cations. Typically 2 Mg^{2+} , 3 Ca^{2+} or abundant monovalent cation, e.g. Na^+ , coordinate phosphate oxygen atoms such that bases face outwards, causing flanking regions of the helix to stack more tightly. Kissing loops arise from base pairing between bases of separate loops; and pseudoknots are hairpin loops which base pair with a complementary region on an ssRNA stretch. Besides non-basepair interactions, RNA readily forms non-canonical base pairs such as Wobble base pairs which

rely on at least one hydrogen bond between bases. These do not support normal Watson-Crick helical structure but causes bases to tilt slightly, yielding a site for hydrophilic molecule binding (at nitrogens and oxygens of the base) and a spacious shallow groove. Both sides often serve as recognition sites (Westhof and Auffinger, 2000). This variety of tertiary structure dictates RNA stability and function.

The secondary structure of RNA can create domains. Domains interact with other domains via tertiary interactions rather than base pairing, and often require metal ions. Besides *cis* interactions (within one molecule of RNA), RNA can associate with other RNAs in *trans* to form multisubunit RNAs (e.g. ribosome), or ribozymes with RNA substrates. It becomes apparent that RNA folding heavily relies on the molecular environment. Presence of cations, hydration state and molecular crowding can affect RNA structural integrity (Butcher and Pyle, 2011).

Eukaryotic mRNAs differ from eubacterial mRNAs in certain ways. Most importantly, eukaryotic mRNAs are post-transcriptionally modified on their 5'-end (RNA capping), on their 3'-end (polyadenylation), and internally (splicing). Also sequence signals can differ. For example, in contrast to eubacterial open reading frames (ORFs), eukaryotic ORF terminator regions include a polyadenylation signal, a sequence which attracts a dedicated polymerase to attach a poly(A) stretch to the 3'-end post-transcriptionally (Zhao, Hyman and Moore, 1999). This poly(A) tail stabilises the transcript by protecting it from 3'-nucleases. In contrast, a prokaryotic poly(A) tail is attached by members of RNA maturation and degradation multienzyme complexes to "tag" a transcript for degradation (Sarkar, 1997).

The 5'-end of a transcript is naturally triphosphorylated, but in eukaryotes mRNAs are modified over several enzymatic steps to attach an inverted methylated nucleotide, most commonly a guanosine, via a triphosphate link. This structure will in this thesis be referred to as "classical eukaryotic cap" and has no bacterial analogue (Ramanathan, Robb and Chan, 2016). The classical eukaryotic cap has multiple roles: transport across the nuclear membrane, translation initiation, and RNA stability (Furuichi, 2015). For a long time, the 5'-RNA cap was thought to distinguish eukaryotic and prokaryotic mRNA. Chapter I of this thesis is concerned with an update of this conception, showing that besides the classic eukaryotic cap, all cells produce a different type of cap, the non-canonical initiating nucleotide (NCIN)-cap. The NCIN-cap is integrated into RNA primary structure during transcription of some RNAs. The potential effect of this newly discovered structure on the function of different RNAs remains to be determined.

1.3 RNA degradation regulates gene expression

Regulation of mRNA stability is a means of regulation of gene expression, and adaption to environmental requirements (Bandyra *et al.*, 2013).

Bacterial RNA degradation differs in Gram positive and Gram negative bacteria (Silva *et al.*, 2011; Mohanty and Kushner, 2016). In *E. coli*, the main break down machinery is the degradosome (figure 1.1). This multiprotein complex comprises of main endoribonuclease RNase E, multifunctional exoribonuclease and polyadenylation enzyme PNPase, RNA helicase Rh1b, and enolase whose function is unclear (Mohanty and Kushner, 2016). Activities of oligoribonuclease and 3'-5'-exoribonucleases is also essential for complete RNA breakdown. A transmembrane domain on RNase E locates the degradosome to the inner cell membrane. Other RNases diffuse freely in the cytosol or are located to foci within the nucleoid. Thus RNA localisation determines the RNA degradation path (Mohanty and Kushner, 2016). Pyrophosphohydrolase RppH was for a while considered the *E. coli* "decapping" enzyme analogous to eukaryotic Dcp2 (responsible for removal of m⁷G cap) (Grudzien-Nogalska and Kiledjian, 2017). Due to the preference of RNase E to monophosphorylated 5'-ends, triphosphorylation is considered protective, and RppH removes this protective molecular structure (Vasilyev and Serganov, 2015). RppH prepares transcripts for degradation by removing 5'-pyrophosphate to generate 5'-monophosphate, the preferred substrate of RNase E (Vasilyev and Serganov, 2015). However, RppH is not a stable component of the degradosome.

RNA processing modulates RNA stability. While eukaryotic poly(A) tails stabilise transcripts, bacterial poly(A) tails function as degradation signals, as they constitute unstructured 3'-ends which are susceptible to exoribonucleases (Régnier and Marujo, 2013). *E. coli* main poly(A)-polymerase PAP1 is not directly associated to the degradosome, but is found in another multiprotein complex with RNA chaperone Hfq, PNPase and RNase E (Mohanty and Kushner, 2016). Hfq is a pleiotropic regulator that is involved in bacterial virulence, motility, biofilm formation and stress response (Fantappiè *et al.*, 2009; Zeng, McNally and Sundin, 2013; Schuergers *et al.*, 2014; Lai *et al.*, 2018). The RNA chaperone pairs trans-acting sRNAs with their target-mRNAs, and has a preference for binding of poly(A) tail structures (Mohanty and Kushner, 2006, 2016; Silva *et al.*, 2011). *Cis*-acting small regulatory RNAs (sRNAs) are transcribed in parallel to their target mRNA from the complementary strand. *E. coli* exonuclease RNase II is inhibited by 3'-secondary structures and requires the unstructured poly(A)-tail as a handle to initiate degradation (Mohanty and Kushner, 2016). Also its

homolog RNase R requires a 3'-ssRNA, and is likely involved in breakdown of Rho-dependently terminated RNAs (Mohanty and Kushner, 2016).

In eukaryotes, mature mRNA is degraded mainly via a deadenylation-dependent pathway. The removal of the unstructured 3'-poly(A) tail precedes other exo- and endonucleolytic steps including 5'-decapping. Alternatively, stability of some individual mRNAs or subclasses of mRNAs can be regulated independently of deadenylation (Beelman and Parker, 1995). For example, within RNA surveillance processes, early nonsense codons lead to 5'-3'-degradation which is initiated by decapping (Beelman and Parker, 1995). Additionally, stability modulating signal sequences can be found throughout the transcript (both in non-translated and translated regions) and interact with RNases or other proteins which then recruit nucleases (Sachs, 1993).

Eukaryotes have an exosome, a multi protein complex with 3'-5'-exonuclease and helicase activity. The exosome is involved in maturation and decay of almost all types of RNAs (transcripts generated by RNA polymerases I, II and III). The 9-protein core of the exosome

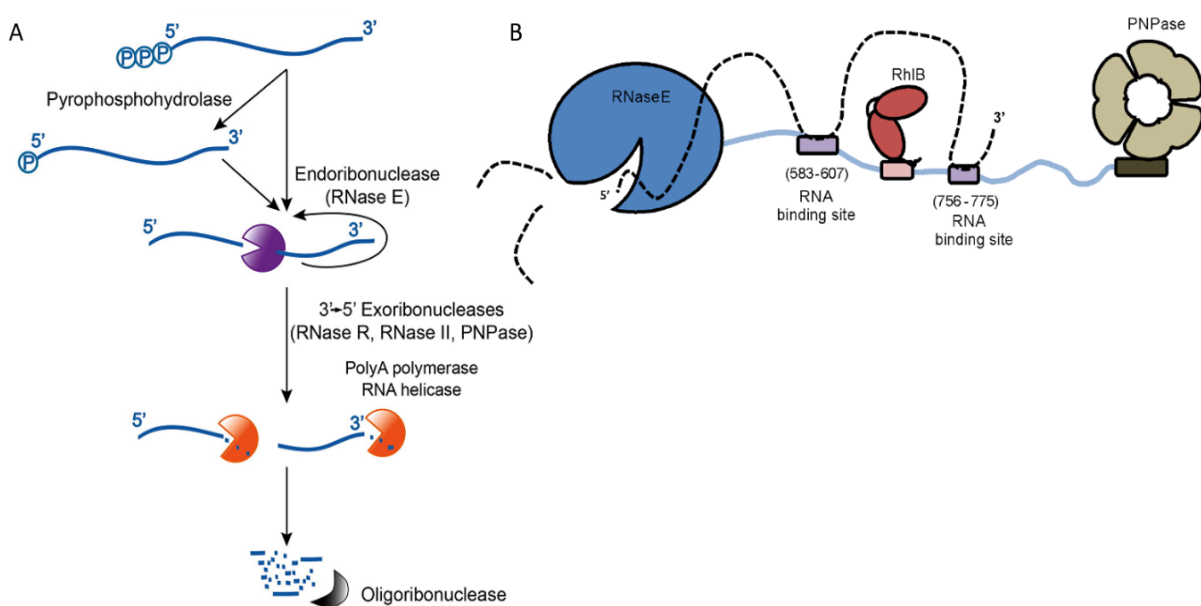


Figure 1.1 The bacterial degradosome (in Gram negatives). A. From (Bandyra *et al.*, 2013). mRNA degradation in *E. coli* starts when 5-triphosphorylated RNA is dephosphorylated by a pyrophosphohydrolase (RppH), which primes it for interaction with endoribonuclease RNase E. RNase E has a preference for 5'-monophosphorylated RNA, although interaction with triphosphorylated RNA is also possible. RNase E also processes polycistronic RNA. Larger RNA fragments are then digested by further 3'-exoribonucleases (RNase R, RNase II and PNPase) and finally oligoribonuclease. RNA helicases and poly(A) polymerase can play a role in making 3'-ends accessible to exonucleases. B. From (Van den Bossche *et al.*, 2016). *Pseudomonas aeruginosa* RNase E interactions. RNase E has several RNA-binding sites and directly interacts with RNA helicase RhlB and PNPase. It further has a membrane-interaction domain which anchors the degradosome to the cell periphery.

forms a barrel-like structure. To both ends of the barrel, an exonuclease binds, Rrp6 and Rrp44. The holoenzyme associates to three protein cofactors which direct different routes of threading of RNA substrate through the barrel using helicase activity. The routes of threading lead to digestion by one of the exoribonucleases. The exosome is predominantly found in the nucleus (Ogami, Chen and Manley, 2018).

The classical eukaryotic m⁷G cap confers resistance to most ribonucleases (Sachs, 1993). It is removed by Dcp2, Nudt3 and Nudt16 NUDIX hydrolases *in vivo* and further by Nudt2, Nudt12, Nudt15, Nudt17 and Nudt19 at least *in vitro* (Grudzien-Nogalska and Kiledjian, 2017). Besides NUDIX hydrolases, in eukaryotes also the DXO family is known to function as decapping enzymes. However, phosphodiesterbond cleavage by DXO removes the complete cap structure (m⁷GpppN-RNA to m⁷GpppN + RNA. Dcp2, Nudt3 and Nudt16 cleave the triphosphate linkage between cap and RNA (Kiledjian, 2018). A DXO-equivalent in prokaryotes has not been found to date. The NUDIX hydrolases and DXO enzymes also play a fundamental role in removal of NCIN-capping, as will be discussed further in chapter I.

Another contributing factor to RNA stability was recently discovered in eukaryotes. The first canonical nucleotide of the transcript, so the nucleotide adjacent to the m⁷G-cap, is usually methylated, giving for example a 2'-O-methyladenosine (Am). But additional methylation giving a N-6,2'-O-dimethyladenosine (m⁶Am) causes resistance of the cap to decapping NUDIX hydrolase Dcp2, thus conferring increased RNA stability as opposed to single methylated initiating nucleotides (Mauer *et al.*, 2017; Kiledjian, 2018). This methylation is reversible and is considered as epigenetic regulation process. 30% of cellular RNA carries the "extended mRNA cap" (Mauer *et al.*, 2017). Other functions of the methylations in eukaryotic mRNA 5'-regions include translation initiation and distinction of own and viral RNA (Mauer *et al.*, 2017; Kiledjian, 2018).

1.4 Many abundant small-molecule metabolites contain a nucleotide-moiety

One important group of biomolecules in the cell are small metabolites which contain a nucleotide part. These metabolites can act as coenzymes of metabolism (e.g. flavin adenine dinucleotide, FAD, nicotinamide adenine dinucleotide, NAD(H) or Coenzyme A, CoA), or as building blocks of macromolecules or intermediates of catabolism (e.g. Uridine diphosphate glucose, UDP-Glc). Some argue that the nucleotide part of these molecules is not required for the cofactor function and thus might be – or once might have been – a handle for interaction

with other, nucleotide-binding molecules (White, 1976). At the example of NAD(H), nicotinamide undergoes redox reactions (NAD^+ gets reduced to NADH, which gets oxidised to NAD^+), while adenosine is irrelevant for the chemical reaction.

Adenine containing cofactors such as NAD^+ have been produced *in vitro* by ribozymatic modification of 5'-ATP-RNA (Huang, Bugg and Yarus, 2000) – a finding that agrees with the hypothesis that they might constitute “relics from an RNA World” (White, 1976; Jeffares, Poole and Penny, 1998). Possibly, in an RNA World, these cofactors might have been covalently attached to ribozymes to catalyse electron transfer reactions. Further theories about the chemical structure of cofactors say that they might have been part of the ribonucleic acid plus protein (RNP). RNP here denotes a hypothetical ancestor molecule. Also, or they might have been preserved in their “dinucleotide”-structure due to co-evolutionary dependence. Co-evolutionary dependence describes the conservation of “frozen accidents”, molecules which might not constitute a Darwinian “fittest” but which are conserved because other molecules have evolved around them and require their continued existence (Jeffares, Poole and Penny, 1998).

Potentially, the findings presented in chapter I and II, and associated research by other groups, might shed new light on the function of these nucleotide handles and explain their evolutionary conservation. Figure 1.2 shows metabolite molecules of interest to this thesis in comparison to adenosine and uridine trinucleotide.

1.5 Transcription is a central cellular process

Transcription is the synthesis of RNA based on DNA sequence performed by DNA-dependent RNA Polymerase RNAP in bacteria and RNAP II in eukaryotes. Transcription is a central step of gene expression and is thus meticulously regulated by the cell, and presents a potent target for bacteriophage-mediated host inhibition, and also drug development.

1.5.1 RNA polymerase structure of bacteria and eukaryotes is similar

If not otherwise stated, the following text refers to bacterial RNAP. RNAP of bacteria and eukaryotes is a multisubunit enzyme composed of several protein subunits (figures 1.3 and 1.4). The bacterial RNAP consists of a minimal set of subunits where the core enzyme is composed of five subunits, $\alpha_2\beta\beta'\omega$. It is complemented by another factor σ and can bind to a variety of transcription factors. The eukaryotic RNAPs are composed of more subunits, but the core has considerable structural similarity to the bacterial core enzyme.

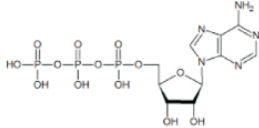
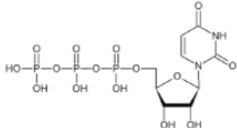
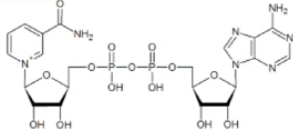
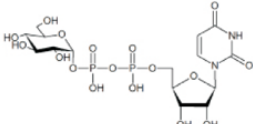
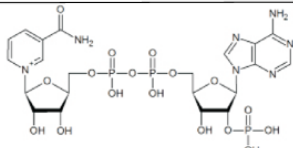
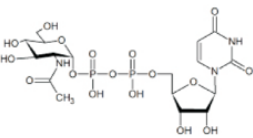
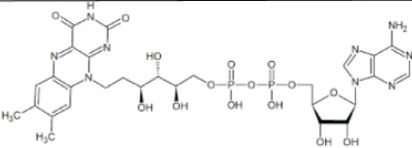
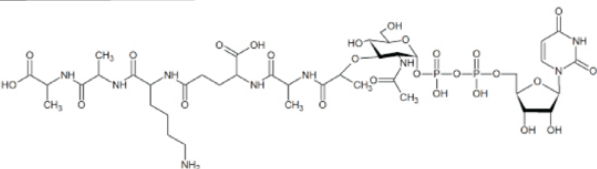
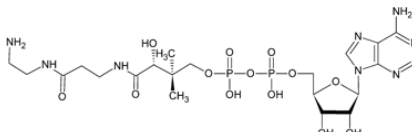
molecular structure	mM*	molecular structure	mM*
 ATP	9.6	 UTP	8.3
 NAD+	2.6	 GlcUDP	2.5
 NADP	0.002	 GlcNAc	9.2
 FAD	0.17	 5AA-MurNAc	0.25**
 DP-CoA	1.4		

Figure 1.2 Molecular structures of ADP-containing molecules ATP, nicotinamide adenine dinucleotide (oxidised, NAD⁺), nicotinamide adenine dinucleotide phosphate (NADP), flavine adenine dinucleotide (FAD) and dephospho-Coenzyme A (DP-CoA) (Julius and Yuzenkova, 2017). UDP-containing molecules are UTP, uridine diphosphate glucose (UDP-Glc), uridine diphosphate N-acetylglucosamine (UDP-GlcNAc) and uridine diphosphate N-acetylmuramic acid-pentapeptide (UDP-MurNAc-5AA). * mM concentration in exponential phase *E. coli* grown on complex medium, from (Bennett, Kimball and Gao, 2009), ** mM concentration in exponential phase *E. coli* grown on complex medium, from (Mengin-Lecreulx, Flouret and van Heijenoort, 1983).

The bacterial core RNAP has a molecular weight of ~ 400 kDa (Zhang *et al.*, 1999; Mathew and Chatterji, 2006). Two copies of α -subunit α I and α II interact with DNA via C-terminal domain, α -CTD. One subunit β and one subunit β' together make up the “crab claw” structure of the core enzyme, formed their coordination towards each other and the DNA template. Further, the ω subunit has mainly structural function, but may also be involved in enzyme regulation and inhibition (Mathew and Chatterji, 2006). Association to σ factor forms the holoenzyme $\alpha_2\beta\beta'\omega\sigma$ and is only necessary for transcription initiation (M. Paget, 2015). The holoenzyme can contain one of several σ for transcription of different sets of genes. *E. coli* has seven sigma factors. Among those are the housekeeping/growth σ^{70} , and the stationary phase σ^S . *M. smegmatis* has 26 sigma factors, including main σ^A and stationary growth σ^B , but also further factors for adaption to the environment, e.g. light induced σ^F (Waagmeester, Thompson and Reyrat, 2005).

A number of functional structural domains have been identified on RNAP (see figure 1.3 B). The catalytic centre is located in a cleft between β and β' (main channel) and binds two divalent metal ions (Mg^{2+}) which function in nucleotide coordination and catalysis of the polymerisation reaction. Mg^{2+} I is held within a cluster of 3 aspartate residues (the triad) of β' (Sosunov *et al.*, 2003). Mg^{2+} II enters the enzyme bound to NTP substrate and is coordinated in the catalytic centre next to Mg^{2+} I. The main channel spans front to back of the enzyme, and is positioned onto DNA by main channel clamp of β' . It accommodates about 12 bp dsDNA (before initiation), or a transcription bubble of about 13 nt, with an 8-9 nt long RNA-DNA-hybrid, during elongation (Nudler *et al.*, 1997; Vassylyev *et al.*, 2002; Bochkareva *et al.*, 2012). σ is partly embedded in the main channel where it will contact DNA with regions 2 and 4 and the catalytic centre with region 3.2 (Borukhov and Nudler, 2008). Bridge helix and trigger loop (TL) protrude into the catalytic site and are essential for catalysis. Folding of TL after binding of correct substrate shifts the active site into the correct conformation for catalysis (Yuzenkova and Zenkin, 2010). The secondary channel between the rear end of the enzyme and the catalytic centre is likely the site of entry for nucleotide substrates (Batada *et al.*, 2004; Vassylyev, Vassylyeva, Zhang, *et al.*, 2007). Lid, zipper and rudder regions serve for structural stabilisation during elongation, by interacting with template DNA or the RNA-DNA hybrid (Gnatt *et al.*, 2001; Kuznedelov *et al.*, 2002; Vassylyev, Vassylyeva, Perederina, *et al.*, 2007; Yuzenkova *et al.*, 2011). RNA chains longer than 8-9 nt are threaded through an RNA exit channel which spans from the main channel through flap, zipper, lid and Zn-finger (Korzheva *et al.*, 2000).

In *E. coli*, RNAP core genes are located within three separate ORFs. Subunit α (gene *rpoA*) is encoded on an operon (alpha operon) together predominantly with ribosomal protein genes, such as *rpsM* (S13) or *rpsD* (S4) (Meek and Hayward, 1984). All genes of the operon except for *rpoA* are regulated by S4, which can repress their translation while *rpoA* translation is unaffected (Meek and Hayward, 1984). β and β' (genes *rpoB* and *rpoC*) are transcribed from a common operon in conjunction to further ribosomal proteins (50S subunit) (Dennis, 1977), and are also regulated independently from the ribosomal protein genes on their operon (Meek and Hayward, 1986). ω subunit gene (*rpoZ*) is transcribed from same operon of SpoT (Gentry and Burgess, 1989), a regulator of stringent response. RpoZ was also hypothesized to execute a secondary function in stringent response that goes beyond being part of transcription machinery (Chatterji *et al.*, 2007). Several loci encode σ , such as *rpoD*, the gene for σ^{70} (Cho *et al.*, 2014).

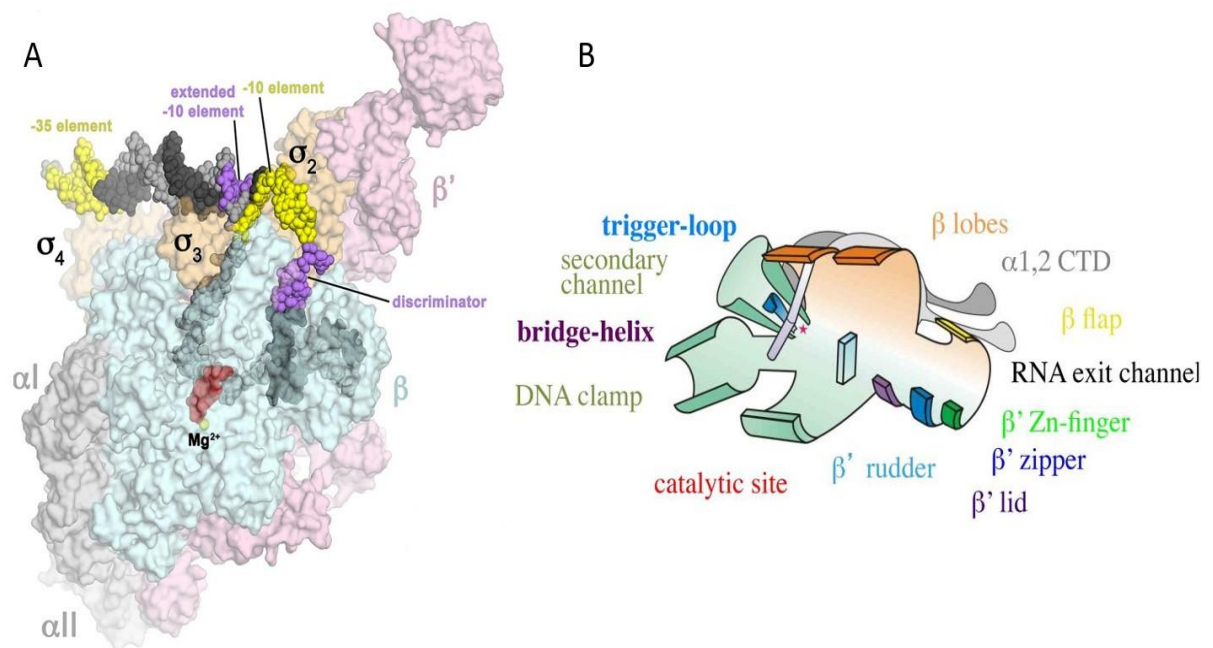


Figure 1.3 Bacterial RNAP subunits and domains. A From (Bae *et al.*, 2015). Structure of RNAP holoenzyme in complex with promoter template strand. Catalytic centre is indicated by Mg^{2+} ion, σ factor (orange) makes contact to DNA (grey with yellow and purple elements) with indicated regions. β subunit is shown in mint green, β' in pink and α in grey. B. From (Nudler, 2009). Important structural regions of the holoenzyme are shown. See text for description.

1.5.2 RNA polymerase catalysis is a two metal ion mechanism

If not otherwise stated, the following text refers to bacterial RNAP. Without σ , RNAP core binds to DNA indiscriminately. *E. coli* σ^{70} domains 4 and 2 recognise consensus sequences around -35 and -10 nucleotides upstream of the transcription start site (TSS, or +1), called -35 element and -10 element, respectively. α -CTD interacts with an upstream AT-rich sequence, the UP-element, which is not present in all promoters. Utilisation of different sigma factors is a form of transcriptional regulation of gene expression in bacteria. Most organisms have several σ factors, to facilitate adaption to different environmental conditions. Alternative σ recognise specialised promoters, thus they direct RNAP towards different genes in alternative conditions. Variations in promoters include variances in recognition site sequences as compared to consensus (consensus in *E. coli* σ^{70} promoters is -35: 5'- TTGACA and -10: 5'- TATAAT), variances in spacer length between those two elements (consensus is 17 +/- 1 nt), absence of UP-element or -35 element, and extension of -10 element by upstream TGn ("extended -10-element"). σ itself can be regulated via anti-sigma factors (which bind σ constitutively until another signal releases σ), or other regulatory proteins. For example, protein Crl activates *E. coli* stress factor σ^{38} to bind to core at higher affinity than without Crl (Gaal *et al.*, 2006). Another growth phase related transcription control pathway in bacteria is via ppGpp in stringent response. The nucleotide analogue and alarmon (p)ppGpp likely acts on transcription by binding protein DksA which then binds RNAP secondary channel and prohibits transcription activity on rRNA promoters (Paul, Barker, *et al.*, 2004), while enhancing transcription on amino acid promoters (Paul, Bergmann and Gourse, 2005).

The catalysis of polymerisation takes place within the active centre of the enzyme (see figure 1.4). Nucleotide binding sites in RNAP are called *i* (initiation site) and *i+1* (next to initiation site). For the first phosphodiester bond formation, 5'-NTP binds *i* site, 3'-NTP binds *i+1*. In elongation, 3'-end of RNA is positioned in *i* site, incoming NTPs in *i+1* site. The sequence of the growing RNA chain is determined by complementary base pairing to the template. Only binding of nucleotides complementary to template facilitates folding of TL to fully close the active site and create the environment for catalysis (Yuzenkova and Zenkin, 2010). Here, NTP substrates and the two metal ions are coordinated to assist nucleophilic attack of 3'-OH of the 3'-NMP of RNA chain on the 5'- α -phosphate of incoming NTP. Mg^{2+} I brings both substrates into the correct position for the reaction to take place. Mg^{2+} II enters the enzyme with incoming nucleotide, bound to the phosphate groups. Mg^{2+} II positioning between pyrophosphate and nucleophilic centre of 3'-RNA facilitates release of pyrophosphate by

hydrolysis of the phosphodiester bond. Subsequently, an ester bond to the hydroxyl group is made (Steitz and Steitz, 1993; Sosunov *et al.*, 2003; Zenkin *et al.*, 2006).

Despite the role of TL in assuring incorporation fidelity, misincorporation occurs at a low rate. This can be corrected by RNAP proofreading activity (figure 1.4 C).

Pyrophosphohydrolysis can be reversed to remove misincorporated nucleotides. The correct substrate is selected at the *E* site of catalytic centre prior to the polymerisation reaction. On the molecular level, matter undergoes oscillations which allow for isomerisation between different stages of molecular interactions. If a mismatch occurs, the 3'-end of RNA has altered flexibility as compared to a correctly paired hybrid, and shifts back into the *E* site. If the 3'-end of RNA disengages from *i* site during pausing, this is called backtracking. Backtracking can be resolved if the 3'-terminal dinucleotide is cleaved to restore correct configuration of the active site, with 3'-terminus resdsing in *i* site (Zenkin, Yuzenkova and Severinov, 2006; Yuzenkova *et al.*, 2010).

Incorporation mistakes, but also specific sequence signals can stall RNAP during elongation. This is called pausing and can lead to termination of transcription, but can also be a means to regulate RNA folding (Artsimovitch and Landick, 2000; Zhang, Palangat and Landick, 2010; Bochkareva *et al.*, 2012). Pausing is also controlled by transcription factors, such as NusA (promotes pausing) or NusG (promotes elongation) (Artsimovitch and Landick, 2000; Herbert *et al.*, 2010). Further, hydrolysis of backtracked RNA can be promoted by transcription factors such as GreA and GreB, which protrude into the active centre to replace TL in positioning of Mg^{2+} II for hydrolysis (Laptenko *et al.*, 2003; Sosunov *et al.*, 2003).

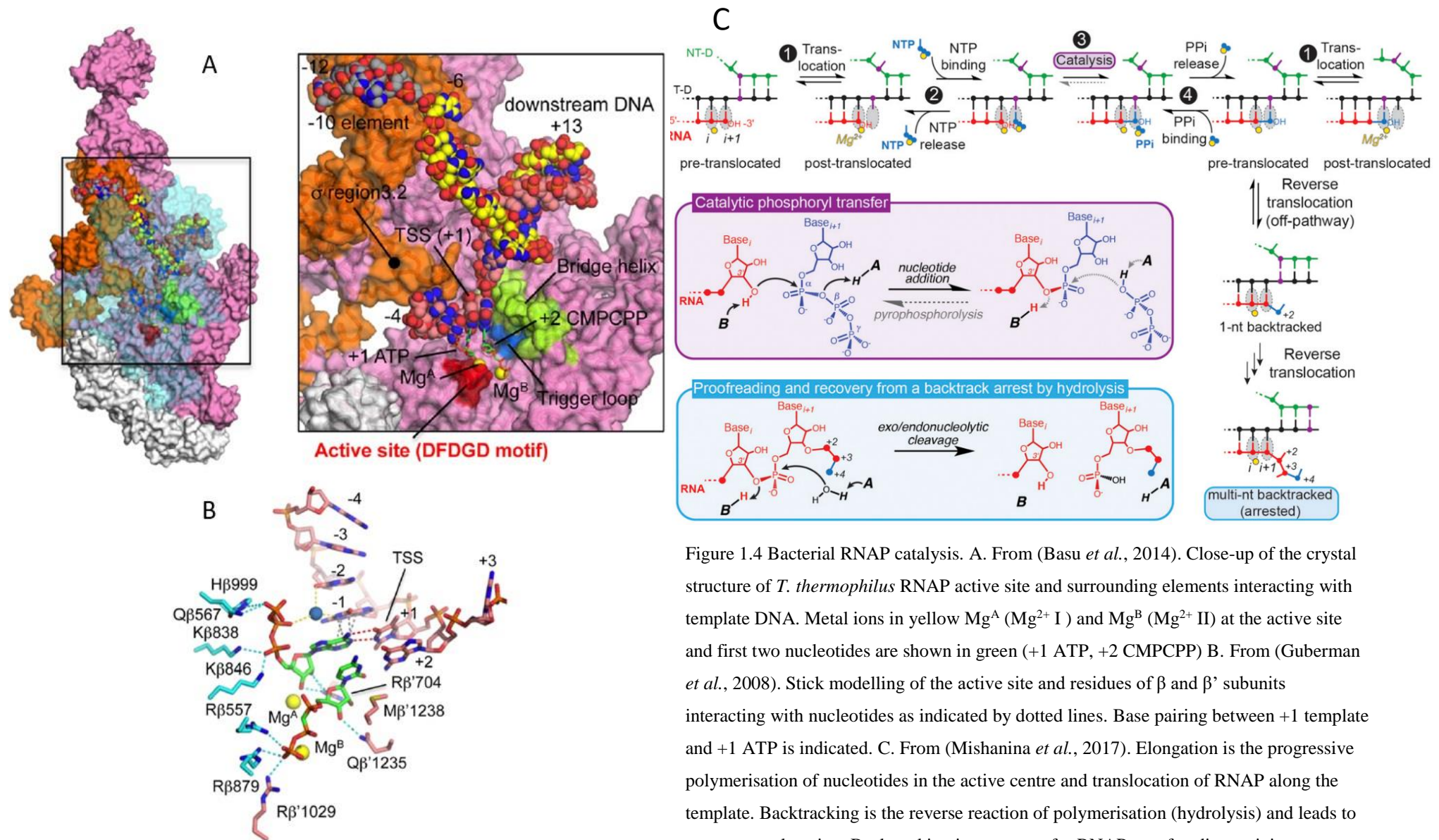


Figure 1.4 Bacterial RNAP catalysis. A. From (Basu *et al.*, 2014). Close-up of the crystal structure of *T. thermophilus* RNAP active site and surrounding elements interacting with template DNA. Metal ions in yellow Mg^A (Mg²⁺ I) and Mg^B (Mg²⁺ II) at the active site and first two nucleotides are shown in green (+1 ATP, +2 CMPCPP) B. From (Guberman *et al.*, 2008). Stick modelling of the active site and residues of β and β' subunits interacting with nucleotides as indicated by dotted lines. Base pairing between +1 template and +1 ATP is indicated. C. From (Mishanina *et al.*, 2017). Elongation is the progressive polymerisation of nucleotides in the active centre and translocation of RNAP along the template. Backtracking is the reverse reaction of polymerisation (hydrolysis) and leads to reverse translocation. Backtracking is necessary for RNAP proofreading activity.

1.5.3 Transcription is a series of distinct steps from initiation over elongation to termination

If not otherwise stated, the following text refers to bacterial RNAP. Transcription initiation can be described in three steps: promoter binding, open complex formation, and abortive initiation. 1) *E. coli* RNAP (EcRNAP) holoenzyme binds to promoter DNA via σ regions 4 and 2 and α -CTD. RNAP spans an area of about 60 nt (-40 to +20) on the template. The RNAP-promoter complex is closed (RP_c), DNA remains double stranded at first. 2) Melting of DNA at the active site via isomerisation of RP_c conformation of RNAP holoenzyme and promoter forms an RNAP-promoter open complex (RP_o). DNA is unwound by about one helix turn (Revyakin *et al.*, 2006). The transcription bubble extends ~15 nt around the TSS. In many organisms, transcription initiation requires additional factors. Mycobacterial RNAP requires RbpA transcription factor on many promoters (Tabib-Salazar *et al.*, 2013; Perumal *et al.*, 2018). Eukaryotic RNAP requires a plethora of transcription factors to initiate transcription. 3) In bacteria, initiation of transcription, or the polymerisation of first few NTPs, takes place frequently before transcription becomes productive. This is referred to as abortive initiation, as strong interaction between RNAP and promoter DNA compete with forwards directed catalysis leading the nascent RNA chain to reverse back into nucleotide binding sites. The RNAP-promoter initial transcribing complex (RP_{init}) produces short abortive RNA products (Revyakin *et al.*, 2006). RNA-DNA hybrids are instable and abortive products are released. Longer RNA products are stabilised in the active centre by base pairing with DNA for about 8 nucleotides (Basu *et al.*, 2014). In contrast, the very first NTP at the *i* site remains unstabilised. Therefore transcription initiation is highly dependent on substrate concentration which at some promoters acts as a regulatory mechanism (Basu *et al.*, 2014). σ region 3.2 ($\sigma_{3.2}$) finger loop protrudes into the catalytic centre such that the growing hybrid clashes into the loop. Growing RNA displaces $\sigma_{3.2}$ loop, and loop pushes RNA backwards into the catalytic centre (Bae *et al.*, 2015).

As soon as a longer transcript (9-11 nt) is produced, RNAP manages to escape from the promoter to form RNA-DNA elongation complex RD_e (also called ternary elongation complex TEC). This was proposed to be a product of downstream DNA stress, caused by scrunching of ssDNA into RNAP (RP_{init} scrunched) (Revyakin *et al.*, 2006). σ factor dissociates. The exact order of events is not known (Basu *et al.*, 2014). While σ is free to bind another core RNAP (or an anti- σ factor), core RNAP progresses to the elongation stage (Bae *et al.*, 2015). Figure 1.5 shows transcription from initiation to elongation.

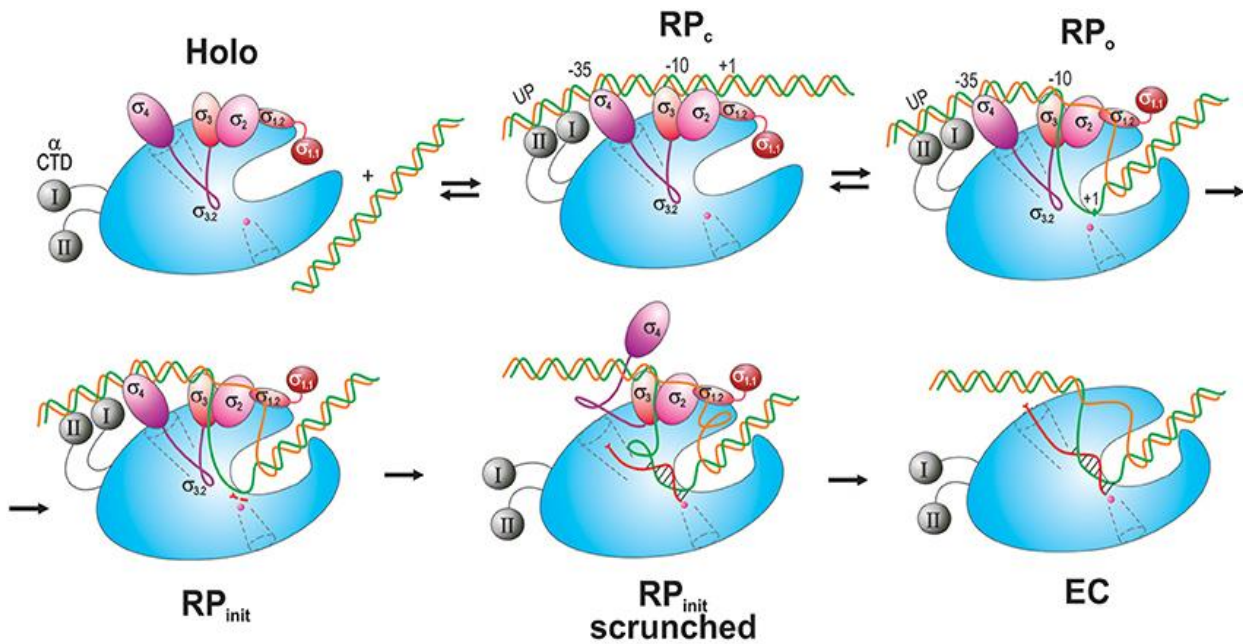


Figure 1.5 From (Lee and Borukhov, 2016). Bacterial transcription initiation and transition into elongation. The RNAP holoenzyme recognises promoter DNA via interactions of α -CTD and σ factor. Holoenzyme isomerisation from RP_c to open RP_o goes along with promoter melting to form the open complex. The first nucleotide is positioned in RNAP catalytic centre (entering through NTP entry channel). A first phosphodiester bond is formed between first and second incoming NTPs by RP_{init} . RP_{init} scrunched struggles forwards as, ssDNA inside of RNAP is scrunched when RNA chain is continued. Downstream DNA stress forces RP_{init} scrunched to escape from the promoter. Nascent RNA is passed through the RNA exit channel, σ is lost and core RNAP proceeds to elongation (EC).

Simply put, elongation is the repetition of phosphodiester bond formation (reviewed in Imashimizu *et al.*, 2014). It requires RNAP moving forwards on the ssDNA template, keeping the transcription bubble open, and NTPs to enter while RNA chain exits. RNAP forwards movement can be explained by two different models. The power-stroke model describes movement as a transition state resulting from coupling of chemical energy into mechanical work. According to this model, a new NTP enters the active site, PPi leaves the active site, and the phosphodiester bond formation is the chemical conversion, followed by RNAP translocation as mechanical work. The Brownian ratchet model disregards any transition states or coupling of chemical reaction and movement. It describes movement as a forwards bias in protein and nucleic acid oscillations caused by cognate NTP occupying and empty active site and being condensed to the RNA chain, followed by another NTP entering and condensation, which is then continued (Imashimizu *et al.*, 2014).

These models can also be applied to explain transcription pausing. In the Brownian ratchet model, fluctuations within the RNA-DNA hybrid correlate to fluctuations in RNAP active site, markedly bridge helix and trigger loop, which are involved in polymerisation and NTP

binding. Some sequences can result in RNA-DNA hybrids of altered flexibility, which disrupts fluctuations of the elongation complex (including RNAP, template and RNA chain), in contrast to smooth translocation. Pausing is overcome by forward energy of phosphodiester bond formation coupled with PPi release, as described in the power-stroke model (Imashimizu *et al.*, 2014). Besides disrupted RNA-DNA hybrid fluctuations, further structural determinants can lead to stalling of RNAP. Road blocking by other proteins, secondary structure formations in nascent RNA chain or transcription factor binding, unusual structure of downstream DNA (e.g. alternative helix shapes), DNA damage (e.g. backbone lesions) in the transcription bubble or mismatched 3'-RNA ends can all lead to TEC stalling (Imashimizu *et al.*, 2014). Backtracking due to mismatch can be resolved by cleavage of a 3'-RNA segment to re-establish active catalytic site conformation. TL works in conjunction with 3'-RNA for autocatalytic cleavage of the transcript (Zenkin, Yuzenkova and Severinov, 2006).

Pausing is a target mechanism for many transcription regulators. For example, promoter-proximal pausing is signalled at 10-20% of *E. coli* transcripts due to a -10 element-like sequence of altered hybrid stability, which needs to be overcome by GreA binding (Marr and Roberts, 2000).

Transcription termination in bacteria can occur via two main pathways. Intrinsic termination (also called Rho-independent termination) is conferred by nucleic acid-RNAP interactions. The nascent RNA chain folds into a hairpin structure followed by a U-rich segment which destabilises TEC. Destabilisation can be caused by hypertranslocation of RNA relative to DNA (hypertranslocation model). Alternatively, the hairpin pulls the 3'-end of the *i* site into the exit channel (slippage model). Or the transcription bubble collapses due to RNAP conformational change triggered by the hairpin loop (allosteric model) (Gusarov and Nudler, 1999; Yarnell and Roberts, 1999; Touloukhonov, Artsimovitch and Landick, 2001; Larson *et al.*, 2008; Santangelo and Artsimovitch, 2011; Molodtsov, Anikin and McAllister, 2014; Roberts, 2019). The main requirement of intrinsic termination is the double stranded segment of the hairpin to be ~8-9 nt in relative distance to the end of the U-rich terminator. *In vitro*, the hairpin can be replaced by DNA or RNA probe binding (Yarnell and Roberts, 1999). Intrinsic termination is related to transcription pausing. Pausing signals are similar to downstream end of intrinsic termination sequences (Gusarov and Nudler, 1999; Yarnell and Roberts, 1999).

In Rho-dependent termination, the ATP-dependent RNA translocase Rho binds to an unstructured C-rich segment on nascent RNA and a conserved region on RNAP and moves in

5'-3'-direction along RNA via ATP-hydrolysis. As it approaches the elongation complex, it pushes RNAP off the transcript (Richardson, 1982; Wu and Platt, 1993; Thomsen *et al.*, 2016). This type of termination was shown to be essential (in organisms who have Rho). Rho activity prevents production of as antisense transcripts, which can lead to DNA damage as they can insert themselves into negatively coiled DNA (Peters *et al.*, 2012; Raghunathan *et al.*, 2018).

A further type of termination is specific to transcription-coupled DNA repair. ATP-dependent DNA translocase Mfd “patrols” DNA via ATP hydrolysis energy until it encounters a transcription elongation complex that was stalled on damaged DNA. It then binds RNAP and removes it from template and also recruits DNA repair enzyme UvrA (Selby and Sancar, 1993; Roberts and Park, 2004; J. Fan *et al.*, 2016; Adebali *et al.*, 2017; Le *et al.*, 2018).

2. Materials and Methods

2.1.1 Supplier lists

Reagents, enzymes and services were purchased from suppliers listed in tables 2.1 and 2.2.

Also reusable consumables are listed, but not one-off use items such as petri dishes.

Manufacturers of devices are given in the text.

Table 2.1 Supplier of chemicals, consumables, kits and commercial media

Supplier	Chemicals, kits, constructs, services
BioRad	Micro Bio-Spin 6
Eurofins	DNA sequencing services
Expedeon	SDS-PAGE precast gels
	SDS-PAGE run buffer 20x
	InstantBlue
Fingerprints Proteomics	Protein Mass Spectrometry
Fischer Scientific	TEMED
Formedium	Glucose
	Kanamycin sulfate
	Carbenicillin
	Amino Acids and Casamino Acids
GE Healthcare	His-Trap FF Ni-column 5 ml
	Resource Q ion exchange column 1 ml
	NTPs
	Gel elution kit
	Streptavidin sepharose
Hartman Analytics	Radiochemicals

Invitrogen	SDS-PAGE NuPAGE precast gels
	NuPAGE Run Buffer
Integrated DNA Technologies (IDT)	DNA oligos, gBlocks
Melford	Proteinase K
Merck-Millipore	0.45 μ M PVDF-filter
	0.22 μ m Whatman filter
Qiagen	all other kits
Remel	<i>M. smegmatis</i> growth media 7H9, 7H10
Santa Cruz	Bovine Serum Albumin
Sigma-Aldrich	all unless mentioned in this list
Spectrum Laboratories	Spectra-Pore 6 Dialysis Membrane
Twist Bioscience	pET28a-p49 construct

Table 2.2 Supplier of enzymes

Supplier	enzyme
Agilent Technologies	QuikChange kit Ultra
Epibio	Terminator enzyme
New England biolabs	all restriction endonucleases
	RNase H + buffer
	DNA Polymerase I + buffer
	T4 DNA Ligase
	NEBuilder HiFi DNA assembly Mix (Gibson Assembly)
	Taq DNA Polymerase
	Q5 High-Fidelity DNA Polymerase
	RppH
	Phusion High-Fidelity DNA Polymerase
	Protein Standard PAGE Ruler
	DNA Standard Gene Ruler, O'Gene Ruler Mix, 100 kb plus
Novagen	Thrombin Cleavage kit
Promega	TSAP Thermosensitive Alkaline Phosphatase
Sigma-Aldrich	Snake Venom Phosphodiesterase

2.1.2 Recipes

Table 2.3 Recipes of homemade media and buffer used in this study

Name	Recipe
23 % Bis:acrylamide Gel	1 x TBE, 7.5 M urea, 20 % acrylamide, 3 % bisacrylamide (0.1 % SPS, 0.1% TEMED)
2x Transcription STOP buffer	1x TBE, 7 M urea, 20 mM EDTA, 100 µg/ml heparin, 0.02 % bromphenol blue, 0.02 % xylene cyanole, in formamide
33 % Bis:acrylamide Gel	0.5 % TBE, 6 M urea, 30 % acrylamide, 3 % bisacrylamide (0.1% SPS, 0.1 %TEMED)
ADC Supplement	50 g/L BSA, 20 g/L D-glucose, 8.5 g/L NaCl
Bis:acrylamide dilution solution	1x TBE, 8 M urea
Dialysis buffer (standard)**	50% glycerol v/v, 20 mM Tris HCl pH 7.4, 200 mM KCl, 1 mM DTT, 0.5 mM EDTA
Grinding buffer	50 mM Tris HCl pH 7.9, 200 mM NaCl, 5% v/v glycerol,
ICP-MS dilution buffer	0.7 mM KCl, 30% polyethylene glycol
ICP-MS internal standard	2 % HNO ₃ , 20 ppb Pt, 20 ppb In
ICP-MS Standards	100 mg/L of all Zn, Mg, Mn, Cu, Co, Ni, Fe in internal standard
Laemmli loading dye	100 mM Tris - HCl pH 6.8, 2% SDS. 20% glycerol, 4% β-mercaptoethanol
LB agar	1% tryptone, 0.5% yeast extract, 1% NaCl, 2% agar

Luria – Bertani (LB) medium	1% tryptone, 0.5% yeast extract, 1% NaCl
<i>M. smegmatis</i> RNAP Dialysis buffer	20 mM HEPES KOH pH 7.9, 50% glycerol, 0.5mM EDTA, 1mM DTT
Middlebrook Top Agar (MBTA)	4.7 g 7H9, 7.0g/L BactoAgar (= 0.4% agar)
MsRNAP transcription buffer	40 mM HEPES KOH pH 7.9, 50 mM NaCl, 5 mM MgCl ₂ , 0.5 mM EDTA, 5 % glycerol
Ni-column buffer A	20 mM Tris HCl pH 7.9, 300 mM NaCl, 5% v/v glycerol, cOmplete EDTA protease cocktail inhibitor
Ni-column buffer B	20 mM Tris HCl pH 7.9, 300 mM NaCl, 5% v/v glycerol, 200 mM imidazole, cOmplete EDTA protease cocktail inhibitor
Ni-column charging buffer	0.2 M NiCl ₂
P49 Dialysis buffer	30 % glycerol, 50 mM K-POH pH 7.5, 3 mM β-mercaptoethanol, 1 mM MnSO ₄
P49 Lysis buffer	50 mM Tris-HCl pH 8, 150 mM NaCl, 3 mM β-mercaptoethanol, cOmplete EDTA protease cocktail inhibitor
P49 Ni-column buffer A	20 mM Tris-HCl pH 8, 250 mM NaCl, 3.5 mM β-mercaptoethanol, 1% triton X-100
P49 Ni-column buffer B	20 mM Tris-HCl pH 8, 250 mM NaCl, 3.5 mM β-mercaptoethanol, 1% triton X-100, 200 mM Imidazole
P49 Priming buffer	50 mM HEPES KOH pH 7.5, 10 mM Mg-acetate, 50 mM K-glutamate, 2 mM DTT
P49 Resource Q buffer A	50 mM Tris- HCl pH 7.5, 1 mM MnSO ₄ , 1 mM MgCl ₂ , 3 mM β-mercaptoethanol
P49 Resource Q buffer B	50 mM Tris- HCl pH 7.5, 1 mM MnSO ₄ , 1 mM MgCl ₂ , 3 mM β-mercaptoethanol, 1 M NaCl
Phage buffer	10 mM Tris-HCl pH 7.5, 10 mM MgSO ₄ ,

Pol I buffer	20 mM Tris-HCl pH 8, 50 mM KCl, 10 mM MgCl ₂
POLMRT transcription buffer	40 mM Tris-HCl pH 7.9, 10 mM MgCl ₂ , 10 mM DTT
Primase buffer	50 mM HEPES KOH pH 7.5, 20 mM Mg-acetate, 100 mM K-glutamate, 10 mM DTT
Resource Q buffer B*	10 mM Tris HCl pH 7.9, 5% v/v glycerol, 0.1 mM DTT, 0.1 mM EDTA, 1 M NaCl
Resource Q buffer A*	10 mM Tris HCl pH 7.9, 5% v/v glycerol, 0.1 mM DTT, 0.1 mM EDTA
SDS-PAGE separating gel 12 %	12% acrylamide, 375 mM Tris-HCl pH 8.8, 0.1 % SDS, (0.1 % SPS, 0.01% TEMED)
SDS-PAGE stacking gel 6 %	6 % acrylamide, 125 mM Tris-HCl pH 6.8, 0.1 % SDS, (0.1 % SPS, 0.01 % TEMED)
SDS-RunBuffer	1x TBE, 0.1 % SDS
Simple Transcription Buffer	20 mM Tris-HCl, pH 7.9, 40 mM KCl (+10 mM MgCl ₂)
Stripping buffer	2 M NaCl, 0.1 M EDTA
TBE buffer	89 mM Tris, 89 mM Boric Acid, 2 mM EDTA pH 8.0
TE buffer	10 mM Tris-HCl pH 8, 1 mM EDTA
Terminator buffer	50 mM Tris-HCl pH 7.9, 20 mM KCl, 20 mM MgCl ₂

*for ion exchange chromatography of NudC and Gp53 EDTA and DTT were omitted

**for dialysis of NudC, EDTA was omitted

2.1.3 Strains, primer and templates

Table 2.4 Strains of bacteria used in this study are listed, genotype and source are given as well.

organism	strain	genotype	plasmid	source
<i>E. coli</i>	T7 Express	<i>fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--TetS)2 [dcm]R(zgb-210::Tn10--TetS) endA1 Δ(mcrC-mrr)114::IS10</i>		NEB
	DH5α	<i>fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17</i>		NEB
	JW4012-AM	F-, Δ(<i>araD-araB</i>)567, Δ <i>lacZ</i> 4787(:: <i>rrnB</i> -3), λ-, <i>rph</i> -1, Δ(<i>rhaD-rhaB</i>)568, Δ <i>qor</i> -745::kan, <i>hsdR</i> 514	pCA24N-dnaB	Dr Matthew Peake
	JW3038	F-, Δ(<i>araD-araB</i>)567, Δ <i>lacZ</i> 4787(:: <i>rrnB</i> -3), λ-, Δ <i>ttdT</i> 755::kan, <i>rph</i> -1, Δ(<i>rhaD-rhaB</i>)568, <i>hsdR</i> 514	pCA24N-dnaG	ASKA strain collection
<i>M. smegmatis</i>	mc1255 wt			Prof. Graham Hatfull
<i>M. smegmatis</i>	mc1255 wt			Prof Dipankar Chatterji
<i>M. smegmatis</i>	mc1255 SM07	His-RpoC		Prof Dipankar Chatterji
Mycobacteriophage D29				Prof. Graham Hatfull

2.1.4 Sequences (vectors, templates and primers)

Table 2.5 Primer (and purpose) used in these studies. F= forward (direction of translation), R= reverse.

Primer name (use)	Sequence
D29 gp53 F NdeI (RE cloning)	GCCCATATGCCATGACAGGCGACATCAACAA GCTC
D29 gp53 R XhoI stop (RE cloning)	ATTCTCGAGTCAAGCAGATTTTGCATAGCACC TCCCTCCCAGG
<i>E. coli acnA</i> F (PCR)	TATGATCAACACAAATATGAAATATTG
<i>E. coli acnA</i> R (PCR)	GGAGCTATGTCTGTCACACC
<i>E. coli dnaB</i> pET insert F (Gibson Assembly)	AGCCATATGGCAGGAAATAAACCTTCAAC
<i>E. coli dnaB</i> pET insert R (Gibson Assembly)	GTGCTCGAGTTATTCGTCGTCGTAAGTGG
<i>E. coli dnaG</i> D309A F (Quik Change)	CAATGTCATTTGCTGTTATGCGGGCGACCGTG CAGGCCGCG
<i>E. coli dnaG</i> D309A R (Quik Change)	CGCGGCCTGCACGGTCGCCCCGCATAACAGCA AATGACATTG

<i>E. coli dnaG</i> D309A R (Quik Change)	CGCGGCCTGCACGGTCGCCCCGCATAACAGCA AATGACATTG
<i>E. coli dnaG</i> D309A R (Quik Change)	CAATGTCATTTGCTGTTATGCGGGCGACCGTG CAGGCCGCG
<i>E. coli dnaG</i> DnaG pET insert R (Gibson Assembly)	GGCAGCAGCCAACTCAGCTTCCTTTCGGGCTC ACTTTTTCGCCAGCTCCTGGTTTAATGT
<i>E. coli dnaG</i> F (Gibson Assembly)	GCGGCCTGGTGCCGCGCGGCAGCCATATGGC TGGACGAATCCCACGCGTATTCATTAATG
<i>E. coli dnaG</i> K229A F (Quik Change)	GCAACGATACCCCCGCATACCTGAACTCGCC GG
<i>E. coli dnaG</i> K229A R (Quik Change)	CCGGCGAGTTCAGGTATGCGGGGGTATCGTT GC
<i>E. coli dnaG</i> K241A F (Quik Change)	CAGACATTTTCCATGCAGGCCGCCAGCTTTAC
<i>E. coli dnaG</i> K241A R (Quik Change)	GTAAAGCTGGCGGCCTGCATGGAAAATGTCT G
<i>E. coli dnaG</i> pET insert F (Gibson Assembly)	GCGGCCTGGTGCCGCGCGGCAGCCATATGGC TGGACGAATCCCACGCGTATTCATTAATG
<i>E. coli dnaG</i> pET vector F (Gibson Assembly)	AGCCCGAAAGGAAGCTGAGTTGGCTGCTGCC

<i>E. coli dnaG</i> pET vector R (Gibson Assembly)	AGCCCGAAAGGAAGCTGAGTTGGCTGCTGCC
<i>E. coli dnaG</i> R (Gibson Assembly)	GGCAGCAGCCAACTCAGCTTCCTTTCGGGGCTC ACTTTTTCGCCAGCTCCTGGTTTAATGT
<i>E. coli dnaG</i> Y230A F (Quik Change)	CGATACCCCCAAAGACCTGAACTCGCCGG
<i>E. coli dnaG</i> Y230A R (Quik Change)	CCGGCGAGTTCAGGTCTTTGGGGGTATCG
<i>E. coli nudC</i> F NcoI no N-term tag (RE cloning)	ATACCATGGATGGATCGTATAATTGAAAAAT TAG
<i>E. coli nudC</i> R NotI for C-term tag (RE cloning)	GTGCGGCCGCTTTCTCATACTCTGCCCCGAC
<i>E. coli rnaI</i> F (PCR)	CAGCAGCCACTGGTAACAGG
<i>E. coli rnaI</i> R (PCR)	GGTTTGTTTGCCGGATCAAGAG
<i>H. sapiens lsp</i> F (PCR)	CATAACCTTATGTATCATACACATACGATTTA G
<i>H. sapiens lsp</i> R (PCR)	GCCGGATAAACTTGTGCTTATTTTTC
L5 P_left for gBlock F (PCR)	GGGACTGTCCAGCGTGACCAGC

L5 P_left for gBlock R (PCR) CAGCTTACCCGATAACCGGGTGGc

L5 P_left genome F (PCR) TGCCGGGACTGTCCAGCGTG

L5 P_left genome R (PCR) ATAACCGGGTGGCTGTCAAACCGGAGAA

L5 P_left mod fusion F (mutagenesis) CAGTAGTGCATTCTTGTGTCAccgcgcgcggcggtt

L5 P_left mod fusion R (mutagenesis) CCTCTCGGATCCGCCTACCGaaccgccgcgc

M. smegmatis rbpA F Nde I (RE cloning) ATTCATATGGCTGATCGTGTCCCTGCGGGGCAG
TCG

M. smegmatis rbpA R stop XhoI (RE cloning) ATTCTCGAGTCAGCTTCCGGTTCCGCGCCGCT
TCC

M. smegmatis rrsA F (PCR) CTGCATGCGGGCTGGCTGGTG

M. smegmatis rrsA R (PCR) GCCGCTGCATTCTCTAGTGGC

M. smegmatis sigA F (RE cloning) AATACATATGGCAGCGACAAAGGCAAG

M. smegmatis sigA R (RE cloning) ACTAAGCTTTCAGTCCAGGTAGTCGCG

pCA24-N MCS F (PCR, Sequencing)	GGCGTATCACGAGGCCCTTTCGTCTTCACC
------------------------------------	--------------------------------

pCA24-N MCS R (PCR, Sequencing)	TTGCATCACCTTCACCCTCTCCACTGACAG
------------------------------------	--------------------------------

pET28a control MCS F (PCR, sequencing)	TAATACGACTCACTATAGGG
---	----------------------

pET28a control MCS R (PCR, sequencing)	CTAGTTATTGCTCAGCGG
---	--------------------

T7 <i>t7a1</i> F (PCR)	GCTCTAATACGACTCACTATAGG
------------------------	-------------------------

<i>t7 t7a1</i> R (PCR)	CACTGGCCGTCGTTTTACAACGTCG
------------------------	---------------------------

pET28a control MCS F (PCR, sequencing)	TAATACGACTCACTATAGGG
---	----------------------

pET28a control MCS R (PCR, sequencing)	CTAGTTATTGCTCAGCGG
---	--------------------

Table 2.6 Templates for transcription and primer synthesis used in these studies

Template	Sequence
DnaG template Priming t -1A	cagACACACACACACTAcaaagc
DnaG template Priming t -1C	cagACACACACACACTCcaaagc
DnaG template Priming t -1T	cagACACACACACACTTcaaagc
DnaG template Priming t wt (-1G)	cagACACACACACACTgcaaagc
DnaG/Poll hairpin template	TTTACGCTTCGTTGACACACACACTGCGCGTTTGGGAAAAC TCTTCCCAAAC TATGATCAACACAAATATGAAATATTGGTCCTGGATGGGCGCGTTTTCTCTGTCGATGCTCTTCTGGGCCGA ACTCCTCTGGATCATTACTCACTGATCCTTGACCCCGCTGCGGCGGGGTGTGCAATTTGCTTTGCCACAAGGTT TCTCCTCTTTTATCAATTTGGGTTGTTATCAAATCGTTACGCGATGTTTGTGTTATCTTTAATATTCACCCTGA <i>E. coli acnA</i> promoter ACAGGATCAGGGCTTCGCAACCCTGTCATTAAGGAGGAGCTATGTCGTCAACC TATGATCAACACAAATATGAAATATTGGTCCTGGATGGGCGCGTTTTCTCTGTCGATGCTCTTCTGGGCCGA ACTCCTCTGGATCATTACTCACTGATCCTTGACCCCGCTGCGGCGGGGTGTGCAATTTGCTTTGCCACAAGGTT TCTCCTCTTTTATCAATTTGGGTTGTTATCAAATCGTTACGCGATGTTTGTGTTATCTTTAATATTCACCCACA <i>E. coli acnA</i> mod promoter ACAGGATCAGGGCTTCGCAACCCTGTCATTAAGGAGGAGCTATGTCGTCAACC ATATCTGCAGCTAGGGCACCAATTTGCGATTAGGGCTTGACAGCCACCCGGCCAGTAGTGCATTCTTGTGTC <i>E. coli rnaI</i> mod promoter 9mer ACCGCGCGCGGCGGTTCCGGTAGGCGGATCCGAGAGGATCGTGCCGGTGCCGGTGAAAATCCGGCGGCAAG

ATTCTCCGGTTTGACAGCCACCCGGTTATCGGGTAAGCTGCAAGCATCACCAACTTGGACCGGTGAGATCAG
GCGGGTAGC

E. coli rnaI wt

CCGGCAAACAAACCACCGCTCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTA
CAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGA
AGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGTTGGTAGCGGTGGT
TTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCTTTTGATCTTTTCTACG
GG

E. coli rnaI 18mer gBlock

CAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGG
CCTAACTACGGCTACACTAGAAGAACTCTCTTTCCTTTCTCCGCTCTGCTAAAGCCAGTTACCTTCGGAAAA
AGAGTTGGTAGCTCTTGATCCGGCAAACAAACC

L5 early lytic promoter P_{left} mod
14mer gBlock

TGCCGGGACTGTCCAGCGTGACCAGCGGATAGCGGCACGTAGTGCCCTTTACAGCCACCGAGAACGCGCCA
TATCTGCAGCTAGGGCACCAATTTGCGATTAGGGCTTGACAGCCACCCGGCCAGTAGTGCAATTCTTGTGTCA
ccgcgcgcgcgcgcttCGGTAGGCGGATCCGAGAGGATCGTGCCGGTGCCGGTGAAAATCCGGCGGCAAGATTCTC
CGGTTTGACAGCCACCCGGTTAT

L5 early lytic promoter P_{left} wt

CGGCACGGGCTGCCGGGACTGTCCAGCGTGACCAGCGGATAGCGGCACGTAGTGCCCTTTACAGCCACCGA
GAACGCGCCATATCTGCAGCTAGGGCACCAATTTGCGATTAGGGCTTGACAGCCACCCGGCCAGTAGTGCA
TTCTTGTGTCACCGCAGCAGCAAGGCGGTAGGCGGATCCGAGAGGATCGTGCCGGTGCCGGTGAAAATCCG
GCGGCAAGATTCTCCGGTTTGACAGCCACCCGGTTATCGGGTAAGCTGCAAGCATCACCAACTTGGACCGG
TGAGATCAGGCGGGTAGC

M. smegmatis 16S rRNA promoter
rrsA wt

CTGCATGAGGGCTGGCTGGTGTGCGGCGTGGCAAGCGCCACATTGCGGGGGTGCGCCGGGTGACCGCGTC
TGACCAGGGAAAATAGCCCTCTGACCTGGGGATTTGACTCCCAGTTTCCAAGGACGTAACCTTATTCCAGGTC

	AGAGCGACACGGCCCAGCCGGGAAGCGAAGACAAAGTCCGAGAGACTCCCCTAAGGTGGGGGATCCTCG CTGCCACTAGAGAATGCAGCGGCTTTTAGCCGCCGGATTCTTGCGCGGCAAAGTCGGGCGTGTTGTTTGAGA ACTCAATAGTGTGTTTGGTGGTTTTTGTGTTGTTGTTTTGTCCGCCTCTTTTCCCGTTTAGGGGTGGATGTT TTTGATGCCAGTTTTGGTGTGTTTTGTTAGGTCAGATTTTCTCTGATTGTGAATTCACCTGTCTTTGGATGGG TTG
T7 promoter t7a1 consensus	GCTCTAATACGACTCACTATAGGGAAAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCGCTAATAACAG GCCTGCTGGTAATCGCAGGCCTTTTTATTTGGATCCAGATCCCGAAAATTTATCAAAAAGAGTATTGACTTA AAGTCTAACCTATAGGATACTTACAGCCATCGAGAGGGACACGGCGAATAGCCATCCCAATCGACACCGGG GTCCGGGATCTGGATCTGGATCGCTAATAACAGGCCTGCTGGTAATCGCAGGCCTTTTTATTTGGATCCCCG GGTACCGAGCTCGAATTCCTGGCCGTCGTTTTACAACGTCG
Human mitochondrial light strand promoter lsp	CATAACCTTATGTATCATACACATACGATTTAGGTGACACTATAGAACTCGAGCAGCTGGATCCAGATAGTA GTATGGGAGTGGGAGGGGAAAATAATGTGTTAGTTTGGGGGGTGACTGTTAAAAGTGCATACCGCCAAAAG ATAAAATTTGAATCTGCACGAGCATCAGCCGTCGCCCTATCCCTTATCTTAACTTTCAAACCACCGTTGAT ATATCCCAATGGCTGCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTAT CCGGC

2.2 Microbiological methods

2.2.1 Bacterial growth and storage

E. coli growth

E. coli doubling time is 20 minutes at optimal conditions (Fossum, Crooke and Skarstad, 2007). *E. coli* were grown at 37°C on solid LB agar or liquid LB medium. This is a complex medium that can be supplemented with antibiotic if required. Strains with plasmid pET28a were supplemented with 50 µg/ml kanamycin sulfate, strains with plasmid pCA24N were supplemented with 34 µg/ml chloramphenicol. Bacterial growth in liquid culture is observed via optical density at 600 nm (OD₆₀₀).

M. smegmatis growth

M. smegmatis mc1255 wt and *M. smegmatis* mc1255 SM07 are resistant to carbenicillin. *M. smegmatis* grows both at 30°C and 37°C, preferably in the dark. Doubling time at optimum conditions is 3 h (Klann *et al.*, 1998). Strains were grown as solid culture on 7H10 agar with 0.5 % glycerol, supplemented with 10% Albumin-Dextrose-Chloride (ADC) supplement and 50 µg/mL carbenicillin. For top agar (pouring of bacterial lawn), cells were mixed Middlebrook top agar (MBTA) (brought to 50°C) supplemented with 10% ADC and 50 µg/mL carbenicillin. Typically, 300 µl stationary phase cells (which corresponds to ~10⁷ cells) are mixed with 3 mL Top agar. At 37°C a bacterial lawn is visible within one to two days, colonies can be picked from 37°C after two days, or at 30°C after 3 days.

M. smegmatis liquid culture was inoculated from a single colony or low volume of smooth pre-culture (1:1000 liquid pre-culture into fresh medium). 7H9 broth was supplemented with 0.25 % glycerol, 10% ADC and 50 µg/mL carbenicillin. Liquid cultures for maintenance contained 0.02-0.04 % tyloxapol to prevent clumping. Liquid cultures for phage infection contained 1-4 mM CaCl₂ (D29 requires 2 mM, TM4 requires 4 mM, Adephagia requires 1 mM). Optionally, 1 mg/ml D-xylose was added to liquid infection cultures to prevent clumping. For protein purification, *M. smegmatis* can be grown without ADC enrichment, to avoid contamination with bovine serum albumin. In that case, instead of 10% ADC supplement, 7H9 broth was supplemented with 1% glucose. Liquid cultures are incubated in the dark under constant shaking (~150 rotations per minute, rpm) *M. smegmatis* grows to stationary phase at 30°C within 3 days, to OD₆₀₀ = 0.2 within 1 day.

Stationary cultures of *M. smegmatis* were kept for several months to inoculate pre-cultures before each experiment.

E. coli / *M. smegmatis* storage

Bacteria can be stored at -80°C in 30% glycerol. For this, 700 µl stationary phase culture was mixed with 300 µl glycerol. Bacteria were left on the bench for 30 minutes before freezing.

2.2.2 *Mycobacteriophage growth, infection and storage*

Our mycobacteriophages were provided by Prof Graham Hatfull, Pittsburgh University, USA. All results shown later represent mycobacteriophage D29.

Table 2.7 Mycobacteriophages provided by Prof Graham Hatfull. Phages, their cluster (phylogenetic relationship) and lifestyle are shown. Phages investigated here in host *M. smegmatis* can also infect *M. tuberculosis*. Presented results in chapter III concern D29.

Phage	Cluster (subcluster)	Infection of <i>M. tuberculosis</i>	Lytic or temperate
Adephagia Δ41Δ43	K (K1)	yes	lytic mutant of Adephagia
D29	A (A2)	yes	lytic
TM4	K (K2)	yes	lytic

Mycobacteriophage plaque picking

This protocol was designed as described in protocols from Actinobacteriophage database (Hatfull, Graham, Russell, Dan, Jacobs-Sera, Debbie, Pope, Welkin, Sivanathan, Viknesh, Tse, 2016). For propagation of a clonal phage culture, phages were picked from a single plaque. For this, a lawn of *M. smegmatis* as described above was mixed with 10 µL of diluted phage suspension before pouring. The MOI was kept around below 1 for plaque picking (e.g. 10^7 cells were infected with 0.5×10^7 phages) thus the dilution depends on a previously identified titre, or several dilutions were plated if the titre was unknown. Dilutions were prepared in phage buffer, supplemented with 1-4 mM CaCl₂.

As soon as plaques were visible, top agar of the plaque was stabbed with a sterile pipette tip. The agar plug was suspended into 100 µL phage buffer and vortexed. A serial dilution was prepared to determine the titre.

Phage titration

This protocol was designed as described in protocols from Actinobacteriophage database (Hatfull, Graham, Russell, Dan, Jacobs-Sera, Debbie, Pope, Welkin, Sivanathan, Viknesh,

Tse, 2016). Phage solution of picked plaque can be titrated via spot test or plate method. These are described in more detail in chapter III. In short, a serial dilution from the starting material in phage buffer was performed ($10^0 - 10^{-9}$). Samples (10 μ l) of dilutions can were spotted on a bacterial lawn (spot test), or intermixed with top agar before pouring the bacterial lawn (plate method). As soon as plaques were visible, they were counted, phage titre was determined as plaque forming units (pfu)/ml with the equation presented in chapter III.

Phage amplification

This protocol was designed as described in protocols from Actinobacteriophage database (Hatfull, Graham, Russell, Dan, Jacobs-Sera, Debbie, Pope, Welkin, Sivanathan, Viknesh, Tse, 2016). To produce phage stock solutions of high titre, a single plaque sample was amplified. For this, information from full plate titration was used to create multiple plates with “lace pattern” (see figure 2.1). The lace pattern depends on plaque size, which is unique to the phage, thus it needs to be determined empirically. As a guideline, MOI 1-10 can lead to lace pattern, thus 10^7 cells could be infected with 10^7 - 10^8 phages. Typically 10 plates of lace pattern were incubated overnight with phage buffer at 4°C standing (3 ml of phage buffer were poured onto one petri dish). The phage buffer was then collected and centrifuged ($10000 \times g$, 15 min) to remove bacterial cells, and then passed through a 0.45 μ m PVDF syringe filter. This phage stock was titrated again.

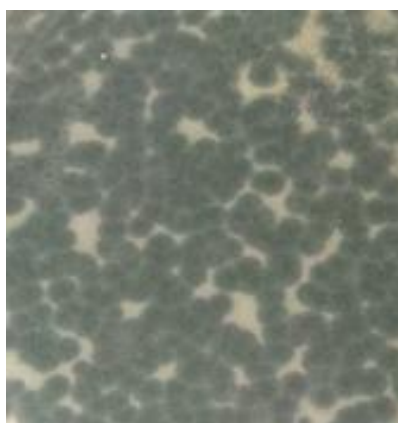


Figure 2.1 D29 plaque lace pattern. Plaque edges blend into each other and little bacterial lawn is left between plaques.

Infection of liquid culture

The method was outlined according to methods described in (Hatfull, 2010a; Parish and Roberts, 2015) and private conversation with Deborah Jacobs-Sera. Liquid pre-culture of *M. smegmatis* was grown in presence of tyloxapol until stationary phase and then used to

inoculate another pre-infection culture. This culture cannot contain tyloxapol, but can contain xylose. A pre-infection culture was harvested at $OD_{600} = 0.2$, early log phase. Only smooth cultures were used. Cell pellets were resuspended in phage buffer supplemented with $CaCl_2$. Phage was added (MOI 10) to the resuspended cells at small volume (synchronisation). The infection was incubated for 30 min at $37^\circ C$ (adsorption). Addition of medium (restoring $OD_{600} = 0.2$) is considered the start of infection ($t = 0$), as growth is facilitated by addition of nutrients.

Phage stocks

Filtered phage suspensions in phage buffer are viable at $4^\circ C$ for several months with some loss of titre. Phage suspensions were renewed (amplified from single plaque) and titrated before use when the stock was more than a month old.

2.3 Molecular Cloning

Polymerase chain reaction (PCR)

PCR from purified DNA was performed using Q5 polymerase, PCR from cell lysates was performed using Phusion polymerase, and colony PCR was performed using Taq polymerase. Table 2. 7 shows general recipe and settings for each polymerase.

After PCR amplification, a sample of the reaction was analysed by agarose gel electrophoresis. For this, a 0.75-1 % agarose gel (in 1x TBE) was poured with Nancy dye and samples were loaded on solidified gel in 1x TBE buffer. Electrophoresis was run at 160 mA. DNA fragments to be used in molecular cloning were purified either from a 1% agarose gel using GE gel elution kit, or directly from the PCR reaction using the PCR reaction clean-up protocol of the GE gel elution kit (if present as a single band). DNA concentration was measured on Nanodrop.

Fragmentation of DNA with restriction endonucleases

Restriction endonuclease recognition sites of pET28a multiple cloning site (MCS) were considered when designing primer (see table 2.5). Further, restriction endonucleases were chosen which are buffer compatible and heat sensitive. Both, vector (~ 1 pmol) and insert (1-2 pmol) were incubated with the restriction endonucleases specified in primer names (table 2.5) at $37^\circ C$ for at least 1 hour, with commercial restriction buffer specified by NEB. Then, thermosensitive alkaline phosphatase (TSAP) was added to the vector reaction according to

manufacturer specifications, and incubation was continued for at least 30 min at 37°C. Then, enzymes were heat inactivated at 74°C for 15 minutes. The reactions were then cooled on ice.

Table 2.8 PCR recipes and PCR machine settings for Q5, Phusion and Taq polymerases

recipe	Q5 polymerase	Phusion polymerase	Taq polymerase
DNA	< 1 µg (0.1 µl purified DNA)	< 250 ng (0.1 µl lysate or 1 colony)	1 colony
dNTPs (each)	200 µM	200 µM	200 µM
Forward primer	0.5 µM	0.5 µM	0.2 µM
Reverse primer	0.5 µM	0.5 µM	0.2 µM
Provided buffer	1x	1x	1x
Polymerase	0.02 U/ µl	0.02 U/ µl	0.025 U/ µl
Final volume	50 µl	50 µl	20 µl
Settings			
Denaturation	98°C / 30 sec	98°C / 30 sec	95°C / 30 sec
DNA melting (loop)	98°C / 10 sec	98°C / 10 sec	95°C / 30 sec
Annealing (loop)	50-72°C / 30 sec	45-72°C / 30 sec	45-68°C / 1 min
Elongation (loop)	72°C / 20 sec per kb	72°C / 20 sec per kb	68°C / 1 min per kb
Final elongation	72°C / 2 min	72°C / 5-10 min	72°C / 5 min
Number of loops	35	35	30

Ligation

For ligation of vector and insert prepared by enzymatic restriction, vector and insert are mixed in a ratio of 1:3 (~0.02 pmol vector and ~0.06 pmol insert) and incubated with 1x T4 DNA ligase buffer and 1 µl T4 DNA ligase/ 20 µl reaction. Incubation was performed 10 min at room temperature or at 4°C overnight. Ligase was then heat inactivated at 65°C for 10 minutes. 2-3 µl of the ligation reaction were transformed into *E. coli* DH5α cells.

Gibson Assembly

Alternative to the molecular cloning protocol using restriction endonucleases, cloning can be performed with Gibson Assembly (NEB, 2015). Here, primers were designed which contain partly overlapping sequences of insert and vector, such that fragments of vector and insert generated by PCR partly overlap and therefore can be directly ligated and transformed. For this, PCR fragments were generated by Q5 polymerase, cleaned up by Gel elution, and mixed in a ratio vector to insert 1:2 (adding up to 0.2 pmol DNA). DNA was mixed with NEBuilder HiFi DNA Assembly Master Mix according to manufacturer specifications and incubated at 50°C for one hour. Afterwards the assembly was cooled on ice and 2 µl were transformed into 50 µl *E. coli* DH5α cells.

QuikChange mutagenesis

To introduce point mutations into a gene on a plasmid, primers were designed according to the Quik Change Site Directed Mutagenesis II manual from Agilent Technologies (Stratagene, 2002). Primer-dimer were complementary to the site to be mutated on DNA sense and anti-sense strand, with the exception of the point mutation. According to *E. coli* codon preference, an amino acid codon was changed by substituting one to two base pairs. Using these primers, the full plasmid containing the gene of interest was amplified with Pfu Ultra polymerase, according to the QuikChange user manual. After the PCR reaction, the assay was cooled briefly to 37°C before addition of DpnI restriction endonuclease for one hour at 37°C. This nuclease digests the parental template (without the mutation). The reaction was stored on ice until transformation into *E. coli* DH5α.

Recombinant genetics (Twist)

Cloning of the human primase catalytic subunit P49 was performed by Twist Bioscience (*Genes - Gene Synthesis / Twist Bioscience*). For this, the P49 cDNA sequence was downloaded (NCBI Reference Sequence: NM_000946.3), and codon optimised for expression in *E. coli*, using the codon optimisation tool of Twist Bioscience. The optimised P49 gene was virtually cloned into the vector pET28a with C-terminal His-tag. A schematic is presented in figure 2.3. The construct can be directly transformed into *E. coli* DH5α.

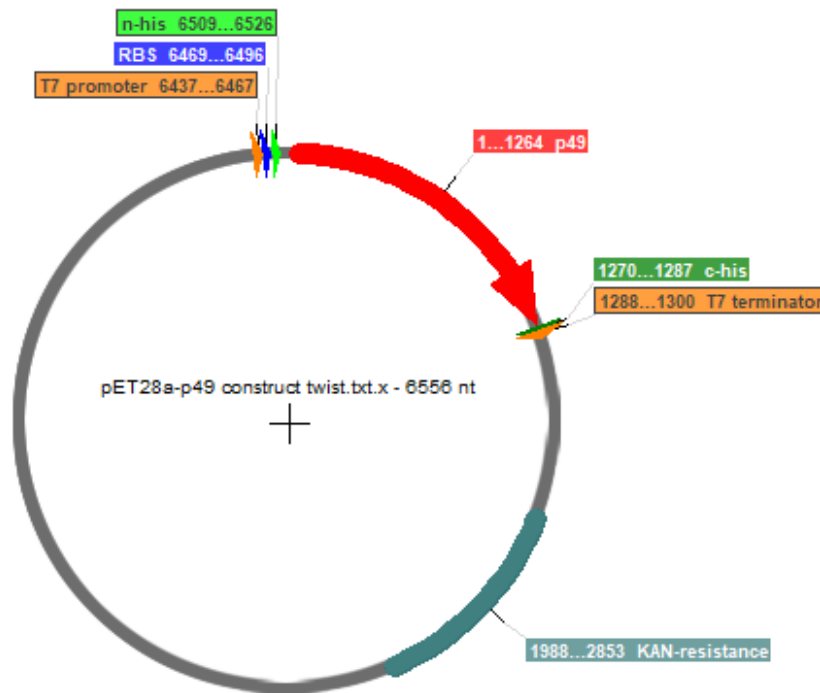


Figure 2.3 pET28a-p49 plasmid as constructed by Twist Bioscience. The gene of P49 is transcribed from the inducible T7 promoter and is fused to a C-terminal His-tag (c-his). The plasmid carries a kanamycin resistance cassette. All cloning for presented studies was performed using pET28a.

Transformation into DH5 α cells

Plasmids were transformed into *E. coli* DH5 α C2987I cells for amplification, according to the High Efficiency Transformation Protocol by NEB (NewEngland Biolabs, 2018) (with modifications). For this, a small amount of DNA (1-1000 pg) was added to 50 μ l competent cells on ice and incubated for at least 30 minutes. The cells were heat shocked for 30 seconds at 42°C. After a short cooling period (5 minutes on ice), transformed cells were mixed with 950 μ L sterile LB medium (room temperature) and incubated shaking (750 rpm on Thermomixer comfort 1.5 ml heatblock by Eppendorf) at 37°C. Then, 100 μ l of a transformation was plated onto LB agar containing 50 μ g/l kanamycin sulfate. Transformants were grown overnight at 37°C. As controls, linearised plasmid (negative control) and circular vector without the gene of interest (positive control) were transformed. 5-15 colonies were subjected to colony PCR to identify transformants that contained the construct. For this, pET28a MCS primer were used to amplify the MCS and distinguish empty MCS from MCS that contains the gene of interest. Up to 5 positive clones were inoculated into 2 ml LB with antibiotic. Plasmids were purified using Qiagen plasmid prep kit (QIAGEN, 2012), in

accordance to the user manual. Samples from plasmid purifications were sent for Sequencing at Eurofins Genomics.

After sequencing of plasmids, constructs which contained the gene of interest without spontaneous mutations and in frame with the desired His-tag, one plasmid clone was transformed into *E. coli* T7 express cells. Transformation into T7 express C2566I cells (NewEngland Biolabs, 2015) was performed according to the above protocol for transformation into DH5 α competent cells. The only modification was that heat shock was performed for 10 seconds and no controls were used. The gained transformants were used for overexpression and protein purification.

2.4 Protein Purification

RNAP Rif-pocket mutants (Q513L, D516V, D516Y, H526P, H526R, H526Y, I572F, and S531I) were kindly provided by Dr Yulia Yuzenkova, Newcastle University, UK. NudC as used in part of the experiments was purified by Dr Amber- Riaz-Bradley, Newcastle University, UK. EcRNAP core and σ^{70} were given by Dr Hamed Mosaei, Newcastle University, UK. Mitochondrial RNA polymerase POLMRT and transcription factors, TFAM, and TFB2M were kindly provided by Dr Andrey Revyakin, University of Leicester, UK.

Purification of recombinant *M. smegmatis* RbpA, recombinant D29 Gp55, and DnaG mutants (K229A, Y230A, K241A, D309A)

Proteins were overexpressed in *E. coli* T7 express cells from plasmid pET28a-*rbpA* or pET28a-*dnaG* (mutants). 1-2 l LB (with kanamycin, 50 μ g/ml or chloramphenicol 34 μ g/ml, respectively) were inoculated using 10-20 ml stationary phase pre-culture (also containing kanamycin/ chloramphenicol). At OD₆₀₀ = 0.5 cells were transferred to a shaking incubator at room temperature and temperature was allowed to adjust until OD₆₀₀ = 0.6. Addition of 1 mM Isopropyl β -D-1-thiogalactopyranoside (IPTG) induces expression of T7 RNAP, which is then able to transcribe the gene of interest from the T7 promoter on pET28a. Cells were grown for 14-20 h and then harvested by centrifugation (7.5 000 \times g, 15 minutes, Beckman JLA 8.1000 rotor). The pellet was stored at -80°C until lysis. Lysis was performed by sonication. *E. coli* cell pellets were topped with ~ 40 ml Grinding buffer supplemented with proteinase inhibitor cocktail. Cells were sonicated (Benson digital sonifier) on ice, at ~30-40 % amplitude 2-3 times for 2 minutes with 2 second pulses and 2 second pulse breaks. The cell debris was removed by centrifugation (Beckman JA25.50 rotor, 10 000 \times g, 15 minutes). The

lysate was filtered through 0.45 μ M polyvinylidene difluoride (PVDF) syringe filter and adjusted to 10 mM imidazole.

Ni-column buffers A and B containing 600 mM NaCl were mixed to give solutions of 10 mM imidazole, 25 mM imidazole and 100 mM imidazole and cooled in ice. A His-Trap Ni²⁺-NTA column with 5 ml column volume (CV) was equilibrated with cold 10 mM Ni-column buffer. Then, the filtrate was passed through the column. The flow-through was collected. The bound protein in the column was washed with 5 CV (25 ml) 25 mM Imidazole Ni-column buffer. The flow-through was collected. Then, bound protein was eluted by 3 CV 100 mM imidazole Ni-column buffer. The eluates were collected in aliquots of 1.5 ml. A final elution of 2 CV 200 mM imidazole Ni-column buffer B was collected.

Afterwards, the His-trap Ni²⁺-NTA column was washed with 5 CV water and then 2 CV 20% ethanol was passed through the column for storage at 4°C. Between purification of different proteins by His-trap Ni²⁺-NTA column, the reusable column was stripped and recharged. For this, the column was washed in 5 CV filtered water, and 4 CV stripping buffer. After another wash step with 5 CV water, the column was recharged with 3 CV Ni-column recharge buffer. Afterwards, column was washed in water again.

Different fractions of the purification process (crude lysate, filtrate, flow-through, wash, elutions) were subjected to SDS-denaturing Polyacrylamide gel electrophoresis (SDS-PAGE) using Expedeon precast gels (4-20% gradient) using Expedion RunBlue SDS-PAGE running buffer. For this, samples were mixed (1:2) with Laemmli dye and boiled for 10 minutes to aid denaturation of proteins. Electrophoresis was performed at 140 mA, until the dye front left the gel. Gels were stained in Instant Blue commercial Coomassie stain.

On the stained gel the protein of interest was identified in reference to its relative size to the protein standard. The purification eluates which contained a visibly pure protein of interest were pooled and concentrated using Amicon Ultra centrifugation filter units. These concentration filters come in various sizes (molecular weight cut-off (MWCO) of 3kDa, 10 kDa, 30 kDa or 100 kDa). The size of MWCO was made in reference to the protein of interest which was then concentrated at 4°C and 5500 $\times g$.

The concentrated protein was dialysed against 1l Dialysis buffer in Spectra/Pore 6 dialysis membrane pre-wetted RC tubing (MWCO 3.5 kDa) at 4°C overnight, and for another 2-3 hours the next day.

Purification of recombinant *M. smegmatis* SigA, recombinant D29 Gp53, DnaB, DnaG wildtype, and NudC,

Purification of protein with this protocol was performed as the previous protocol with some modifications.

Proteins were overexpressed from pET28a (kanamycin resistance, genes *sigA*, *gp53*, *nudC*) or pCN24A (chloramphenicol resistance, genes *dnaG*, *dnaB*). Gp53 was overexpressed at room temperature for 3 hours. SigA induction was commenced at a cell density of OD₆₀₀ = 1.0.

After Ni²⁺-affinity chromatography (buffers containing 600 mM NaCl), proteins were found to be visibly impure. Therefore, they were subjected to ion exchange chromatography. For this, eluates were pooled and diluted with Resource Q buffer A to yield a salt concentration of less than 100 mM NaCl. Buffers and proteins were kept on ice at all times. A 1 mL Resource Q anion exchange column was connected to the AKTA Explorer FPLC from GE Healthcare. The pumping system was equilibrated to Resource Q buffer A, before the protein was loaded via P960 sample pump at a flow rate of 2 ml/min. The flow-through was collected. The column was washed in 0% B buffer. The protein was eluted from the ion exchange column by gradual increment of buffer B (raising salt concentration from 0 to 70% B (700 mM NaCl) within 50 min at a flow rate of 1 ml/min. Elutions were collected guided by the UV Absorbance at 260 nm. At 70% B, the elution gradient was changed to an increment 70-100% B within 10 minutes. The column was washed in 100 % B for 10 column volumes, and then washed with water for at least 10 CV before storage was facilitated by washing in 5-10 CV 20% ethanol.

The protein was identified after SDS-PAGE on NuPAGE precast gels (4-12 % gradient) with commercial NuPAGE run buffer. Visibly pure protein eluates were pooled and concentrated and dialysed as described above.

Purification of recombinant human P49

Purification of P49 was performed at constant supplementation of MgCl₂. *E. coli* T7 express cells with pET28a-p49 were grown in 2l LB with 50 µg/ml kanamycin sulfate and 1 mM MgCl₂ to OD₆₀₀ = 1.6 at 37°C. The culture was then transferred to a room temperature shaking incubator and overexpression was induced with 1 mM IPTG for 3 h. Lysates were sonicated in P49 lysis buffer. A first affinity purification with a 5 ml His-trap Ni²⁺-NTA column was performed with P49 Ni-column buffer. Pooled eluates were dialysed overnight against P49 resource Q buffer A, supplemented with 80 mM NaCl at 4°C. Protein was loaded

onto a 1 ml Resource Q column in AKTA Explorer FPLC equilibrated to P49 ion exchange buffer A, and eluted by increment of buffer B. After SDS-PAGE on NuPAGE gel, visibly pure eluates were pooled and dialysed against P49 dialysis buffer. After dialysis, different P49 aliquots were stored at 4°, -20°C and -80°C. A first experiment was performed immediately after dialysis in case activity would be lost soon. Activity of the protein was equal for all storage conditions on the first days after purification. Only P49 stored at -80°C was used repeatedly the first day.

Purification of *M. smegmatis* His-RNAP

For isolation of His-tagged RNAP from *M. smegmatis*, 10 l *M. smegmatis* SM07 liquid culture (7H9 supplemented with 0.25 % glycerol, 1 % glucose, 0.04 % tyloxapol and 50 µg/ml carbenicillin) were grown to OD₆₀₀ = 0.8. Then, the culture was pelleted by centrifugation (10000 × g, 30 minutes, Beckman JLA 8.1000 rotor), and the pellet was frozen (-80°C). The pellets were topped with about 40 ml Grinding buffer with proteinase inhibitor. The Benson digital sonifier was set to an amplitude of 18-20 % with 15 sec pulses and 15 sec breaks for 5 minutes. The lysate was spun down (15 000 × g, 10 min, 4°C) and filtered. Ni-column purification was performed as described above, with Ni-column buffers containing 300 mM NaCl. Afterwards, Resource Q ion exchange chromatography was performed as described before. Pooled fractions were concentrated to about 200 µl and dialysed against *M. smegmatis* dialysis buffer using Pur-A-Lyzer Mini 12000 dialysis tubes.

Purification of mycobacteriophage-infected *M. smegmatis*

For purification of RNAP from infected *M. smegmatis* (and as control non-infected or non-His-tagged RpoC), purification was done at a smaller scale.

To test the method, 200 ml *M. smegmatis* SM07 were grown to OD₆₀₀ = 0.2. Cells were pelleted and frozen and then lysed by a) Zirconia BeadBeating or b) Sonication.

- a) Zirconia glass beads (0.1 mm diameter) were washed three times in Grinding buffer. Washed beads were transferred into the BeadBeater vial alongside the frozen pellet, to yield a volume ~ 70 ml. Using an ethanol-ice mixture, the BeadBeater was cooled while the cells were lysed with 20 sec pulses, interrupted by 1 min breaks five times. The lysate was recovered and centrifuged (15 000 × g, 15 min, 4°C) and then filtered through a 0.45 µm low protein binding PVDF syringe filter.
- b) Sonication was performed as described above, but in lower volume (7 ml Grinding buffer).

Filtered lysates were adjusted to 10 mM imidazole. 500 µl of Ni²⁺-NTA agarose beads (corresponding to one CV) were equilibrated to 10 mM imidazole Ni-column buffer and added to the lysate and incubated 20 min on ice. The supernatant was recovered. Beads were washed in 1 CV 25 mM imidazole Ni-column buffer three times and the washing supernatant was kept. Protein was eluted from beads by adding 200 µl 100 mM imidazole Ni-column buffer twice and 200 µl 200 mM imidazole Ni-column buffer once. RNAP-containing elutions were identified by SDS-PAGE and protein was concentrated and dialysed against *M. smegmatis* dialysis buffer using Pur-A-Lyzer Mini 12000 dialysis tubes.

For infection, 300 mL *M. smegmatis* SM07 were grown in 7H9 broth supplemented with 0.25 % glycerol, 1 % glucose, 1-4 mM CaCl₂ and 50 µg/ml carbenicillin to OD₆₀₀ = 0.2. Cells were infected with mycobacteriophage at MOI 10 (as explained in chapter III) and after adsorption, culture volume was restored with fresh medium. After predetermined time points (for D29 15 minutes and 40 minutes after infection), cells were pelleted by filtration. For this, the culture was passed through a 0.22 µm Whatman filter using a vacuum pump (Laboport, KNF Lab). Retained cells were frozen on the filter membrane (-80°C). For lysis, cells were washed off the filter using 7 ml Grinding buffer and lysis was performed via sonication as described above. Ni-column buffers for purification of RNAP from infected cells had a salt concentration of 150 mM NaCl. Ni²⁺-NTA agarose bead purification of infected lysates and controls was performed as described above. However, after concentration of eluates (to ~ 50 µl), proteins were loaded onto a self-made SDS-PAGE gel (12 %), excised from the gel and sent to the Proteomics facility (as described in chapter III).

Thrombin cleavage for removal of His-tag

His-tag was removed from Gp53 and RbpA by using the thrombin cleavage protocol by Novagen (*Thrombin Cleavage Capture Kit*, no date). Biotinylated thrombin enzyme cleaves a specific amino acid sequence that is positioned on the peptide between His-tag and gene product. 10 µg of His-tagged protein was digested by 10 µU of thrombin in presence of 10 µM heparin. The final volume including provided thrombin cleavage buffer was 500 µl. After incubation overnight at 4°C while shaking, 50 µl streptavidin beads (provided in the Thrombin cleavage kit) were added to remove thrombin enzyme and the mixture was incubated shaking at 4°C for another 30 minutes. After centrifugation (500 × g, 5 minutes, room temperature), supernatant was recovered and adjusted to 10 mM imidazole. 50 µl Ni²⁺-NTA agarose beads (equilibrated in Ni-column buffer, 10 mM imidazole) were added. The bead supernatant was recovered again after 5 minutes shaking incubation at room temperature

and centrifugation to pellet the beads. By this protocol, about 55 % of the added gene product were recovered without His-tag.

2.5 NudC metal analysis

Inductively coupled plasma mass spectrometry (ICP-MS) was performed by Dr Kevin Waldron, Newcastle University, UK. Metal stock solutions (Pt^{2+} , In^{3+} , Zn^{2+} , Mg^{2+} , Mn^{2+} , Cu^{2+} , Co^{2+} , Ni^{2+} , Fe^{2+} , all at 1000 ppm = 1000 mg/l) were provided by Dr Kevin Waldron. First, an internal standard containing 2 % HNO_3 and 20 ppb Pt^{2+} and In^{3+} ions was mixed. A series of metal standards was prepared alongside different samples. Metal stocks at high concentration were diluted in internal standard to 100 ppm (or 100 mg/l). 100 ppm metal ions were mixed to create a standard series of 0 (only internal standard), 1, 5, 10, 50, 100, and 500 $\mu\text{g/l}$ of each metal in the mixture. Of “red” NudC (purification from overexpression in complex medium, purification in absence of EDTA and DTT) and “white” NudC (red sample was dialysed in dialysis buffer containing 10 mM EDTA until colour was not visible by eye) were adjusted to 100 μM protein concentration. 100 μl of this were mixed with 400 μl internal standard and then one volume (500 μl) of 70 % HNO_3 and heated to 70°C for 10 minutes. 100 μl of this were mixed with 1 ml standard series and volumes were adjusted to 5 ml by adding internal standard. Thus, of seven samples per protein, one sample contains only internal standard and no other metals. Additionally, a control series without protein was prepared. 5 ml samples were mixed with one volume of 70 % HNO_3 . The final samples of 10 ml were predicted to contain 1 μM NudC.

2.6 In vitro RNA synthesis

2.6.1 Synthesis of transcription templates

For transcription assays we used linear, PCR-generated dsDNA fragments which contained a promoter sequence.

Templates *rnai* (promoter of *E. coli* sRNA RNA1) and *rrsA* (promoter of *M. smegmatis* 16 s rRNA) were generated by PCR from heat-lysed *E. coli* or *M. smegmatis*. P_left wildtype template was generated using heat-lysed L5 particle supernatant. Templates *t7a1* consensus, *acnA*, *acnA* mod, *rnai* mod and *lsp* were provided by Dr Yulia Yuzenkova and PCR amplified. P_left 14mer and *rnai* 18 mer templates were ordered as gBlocks from IDT and amplified by PCR.

PCR products were purified from a 1 % agarose gel using the gel elution kit from GE Healthcare.

Templates for primase synthesis were ordered as ssDNA oligos from IDT.

2.6.2 *In vitro* transcription with *E. coli* RNA polymerase from *E. coli* or T7 phage promoters

In vitro transcription experiments were performed as described in (Julius and Yuzenkova, 2017). For initiation complex formation, 0.3 pmol RNAP core enzyme (wildtype or rif-pocket mutant) and 1 pmol σ^{70} , $\sigma^{70\Delta 3.2}$, or σ^S . The promoter template *t7a1* consensus, *rnaI* or *acnA* was provided at 2 pmol. The mixture was incubated in transcription buffer without MgCl₂ at 37°C for 10 minutes. DTT in the assay was carried over from protein storage buffer to a final concentration of around 0.2 mM DTT. After addition of 500 μ M of initiator (ATP, NAD⁺, NADH, NADP⁺, DP-CoA, FAD UTP, UDP-Glc, UDP-GlcNAc, or UDP-MurNAc-AA5), unless otherwise stated, also 50 μ M (12.5 Ci/mmol) of the +2 nucleotide to be incorporated, α (P³²)-CTP (on *rnaI* and *acnA* promoters) or α (P³²)-UTP (for *t7a1* promoter) were provided. For abortive initiation, reactions were started by adding 10 mM MgCl₂ and stopped after 3 minutes by adding 1 volume of transcription STOP buffer.

For synthesis of 9mer products from *rnaI* mod or *acnA* mod RNAP-promoter complexes, 20 μ M UTP and 20 μ M α (P³²)-CTP (20 Ci/mmol), or 25 μ M ATP and 25 μ M α (P³²)-CTP (25 Ci/mmol), respectively, were added. Rifampicin was given at 5 μ g/ml or omitted. These reactions were started by addition of Mg²⁺ containing buffer and stopped after 10 minutes of incubation at 37°C.

For transcription from *rnaI* 18 template, first 1 mM NAD⁺ and 100 μ M UTP, 10 μ M and 25 μ M of α (P³²)-CTP (12.5 Ci/mmol) were added to RNAP-promoter complexes in transcription buffer in a volume of 20 μ l. The reaction was initiated by addition of 10 mM MgCl₂, and then incubated at 37°C for 10 minutes unless otherwise stated. Then, 20 μ l Ni²⁺-NTA beads were added to the reaction, bead supernatant was discarded and beads were washed three times in transcription buffer without MgCl₂. Then, 10 μ M of each CTP, GTP and UTP were added to produce 23mer RNA. The reaction was adjusted to 20 μ l by addition of MgCl₂-containing transcription buffer. RNA synthesis was stopped after another 2 minutes incubation at 37°C.

2.6.3 NCIN-decapping with NudC

For decapping of 9mer RNA produced by NCIN with EcRNAP on *rnaI* mod, 1 pmol RNAP and 1 pmol σ^{70} were mixed with 1 mM ATP, NAD⁺, or FAD and 0.5 pmol *rnaI* mod template.

The reaction was started by adding 25 μM $\alpha(\text{P}^{32})\text{-CTP}$ (12.5 Ci/mmol), and 100 μM UTP in transcription buffer with MgCl_2 . The final volume of this reaction was 10 μl . After 10 minutes at 37°C, 20 μl Ni^{2+} -NTA agarose beads were added. Beads were incubated another 5 minutes. Then, the supernatant was retrieved and split into two assays. To one, 10 pmol NudC were added, to the other the equivalent volume of dialysis buffer was added. Decapping and control reactions were incubated at 37°C for 10 minutes before the reaction was stopped.

Decapping of 18 mer and 23 mer RNA was performed after transcription from *rnai* 18 template. Here, an RNAP-promoter complex was preassembled by mixing 5 pmol RNAP and 10 pmol σ^{70} with 5 pmol DNA in a final volume of 10 μl . To 5'-biotinylated *rnai* 18 template 20 μl streptavidin beads were added. After 5 minutes incubation at room temperature, the reaction volume was adjusted to 20 μl by removing bead supernatant. Onto 20 μl RNA-promoter complex-loaded beads 10 μl reaction mixture containing transcription buffer with MgCl_2 , 1 mM NAD^+ , 100 μM UTP and 50 μM $\alpha(\text{P}^{32})\text{-CTP}$ (25 Ci/mmol) were added. After 5 minutes of incubation at 37°C, the beads were washed twice in transcription buffer (no MgCl_2), containing 1 M NaCl, and three times in transcription buffer (no MgCl_2). Then the original volume (30 μl) was restored and 20 μl of the mixture were removed and split into two samples. To the residual beads, 10 μl transcription buffer with MgCl_2 , containing 10 μM GTP and 10 μM CTP were added. The 23mer reaction was incubated for 5 minutes and washed as before, the original volume was restored (20 μl) and split in two. To 10 μl of 18mer or 23mer RNA, 10 pmol NudC or an equal volume of protein storage buffer were added. NudC reaction was initiated by adding 10 mM MgCl_2 and incubated for 10 minutes at 37°C. Reactions were stopped by adding one volume of STOP buffer.

2.6.4 NCIN-capping by POLMRT and decapping by NudC

The experiment was performed as described in (Julius, Riaz-Bradley and Yuzenkova, 2018). 50 nM POLMRT were mixed with equimolar concentration of transcription factors TFAM and TFB2M and *lsp* promoter DNA. 1 mM of adenosine-containing initiator were added (or as control no additional initiator). Then, the reaction was started by providing POLMRT transcription buffer (with MgCl_2), 50 μM ATP, 300 μM GTP, and 10 μM $\alpha(\text{P}^{32})\text{-UTP}$ (25 Ci/mmol). The reaction was incubated for 30 min at 37°C. Then, 0.5 pmol NudC was added to half of the reactions, or an equal amount of protein storage buffer. Incubation was continued for another 15 minutes before reactions were stopped by adding transcription STOP buffer.

2.6.5 *In vitro* transcription with *M. smegmatis* RNA polymerase from *M. smegmatis* or L5 phage promoters

Transcription experiments were performed in 10 μ l volumes using 0.5 pmol MsRNAP and optionally, 1.5 pmol σ^A , RbpA or Gp53, or any combination of these. DNA (*rrsA*, P_left or P_left mod) was provided at 2 pmol. Unless otherwise indicated, RNAP was mixed with other proteins first, then DNA was added and then the reaction was started by adding *M. smegmatis* transcription buffer containing MgCl₂ and NTPs. Non-radioactive UTPs were added at 100 μ M, or 10 μ M if the same NTP was added as radiolabelled. α (P³²)-UTP (for *rrsA*) or α (P³²)-CTP (for P_left) were added at 25 μ M (12.5 Ci/mmol). Only for order of addition experiment (figure 5.13), CpA initiating dinucleotide was added at 100 μ M and other nucleotides at 50 μ M to yield a 15 nt transcript on template P_left mod. To stop a reaction, one volume of transcription stop buffer was added to the assays. For visualisation of the reaction at different time points, 3 μ l STOP buffer were prepared and 3 μ l were removed from the reaction which was incubating at 37°C and added onto STOP buffer.

When EcRNAP was used on *M. smegmatis* promoters, 0.5 pmol EcRNAP core and 1.5 pmol σ^{70} were optionally mixed with 1.5 pmol Gp53 or Gp55.

2.6.6 *In vitro* primer synthesis by DnaG

For synthesis of primer RNA from templates “priming t” or “hairpin template”, 1 μ M DnaG were mixed with 1 μ M template DNA and 1 mM initiator, unless otherwise indicated. If DnaB was present (3 μ M), DnaG and DNA were premixed and then added to DnaG premixed with initiator. Primer synthesis was initiated by addition of 10 μ M α (P³²)-GTP (5 Ci/mmol), 100 μ M UTP, and primase buffer containing Mg²⁺. The reaction was incubated at 30°C for 10 minutes. The reaction was stopped by adding transcription stop buffer, or salt to a final concentration of 1 mM NaCl.

Kinetics experiments were performed as described above, but α (P³²)-UTP was used to label primer RNA, thus 10 μ M UTP and 100 μ M GTP were given. A large reaction was started and for each time point a small volume was added onto STOP buffer.

2.6.7 NCIS versus ppGpp competition assay

To test the inhibition of primer synthesis by ppGpp, we incubated several concentrations of ppGpp (0, 100 μ M, 600 μ M and 1 mM) with several concentrations of initiators ATP (10 μ M, 60 μ M, and 100 μ M), NADH and FAD (both 100 μ M, 600 μ M and 1 mM), giving 12 samples per initiator. 1 μ M DnaG was pre-mixed with 1 μ M priming t. ppGpp/initiator mixes

were added. After that, reactions were started and radiolabelled as described above, incubated at 30°C for 10 minutes, and then stopped by addition of transcription STOP buffer.

2.6.8 Primer decapping by NudC

To test whether NudC can cleave 5'-NADH on primer annealed to DNA, DnaG was removed after primer synthesis. For this, the reaction was stopped by adding 1 mM NaCl. Then, Ni²⁺-NTA agarose beads (in primase buffer containing 1 mM NaCl) were added and briefly incubated at room temperature. Bead supernatant was passed through a Micro-Bio Spin 6 gel filtration column equilibrated to transcription buffer with Mg²⁺. The filtrate was split into two reactions, one containing 30 pmol NudC, and a control without NudC. Reactions were incubated at 37°C for 10 minutes.

2.6.9 Primer cleavage by RNase H, terminator, SVP, or RppH

For cleavage of primer by other enzymes, shown in figure 4.12, DnaG was removed from the priming template and the RNA-DNA hybrid was filtered as described above, but the Micro-Bio Spin 6 gel filtration column was equilibrated in water. Reactions were mixed with 40 pmol NudC, 1.25 U RppH, 2.5 U Pol I, or 15 pmol SVP with transcription buffer (including Mg²⁺). 1.25 U RNaseH was added with commercial RNase H buffer. The final reaction volumes were equal and contained 1x buffer, to the control only water was added. Reactions were incubated 10 minutes at 37°C before STOP buffer was added.

2.6.10 Primer removal by Pol I

For investigation of Pol I activity on NCIS-primer, primer were synthesised by DnaG on hairpin template, as described above, except for the concentration of ATP. While other initiators were provided at 1 mM, 100 µM ATP were given due to difference in K_M to get a comparable amount of primer substrate for Pol I. DnaG was removed by addition of salt and Ni²⁺-NTA beads. The hybrid was filtered and eluted in water. A sample (4.2 µl) was retained and mixed with water (0.8 µl) as t = 0. 10 U of Pol I were premixed with 500 pmol of dNTPs and 3 µl 10x Pol I buffer to a volume of 5 µl. This was added to the reaction to a final volume of 30 µl. 5 µl of this were removed to stop the reaction at different time points.

2.6.11 Primer initiation by P49

For primer initiation by P49, 2 µM P49 were mixed with 8 µM ssDNA template and 1 mM of initiators guanosine, GMP, GTP, dGTP, m⁷G, GpU, or ATP. 50 µM α(P³²)-UTP (25 Ci/mmol), and 10 µM GTP in P49 priming buffer. This was incubated at 37°C for 10 minutes. The reaction was stopped by adding STOP buffer.

2.6.12 Denaturing bis:acrylamide gel electrophoresis

Radiolabelled RNA products of transcription or primer synthesis were separated by size and charge by electrophoresis on homemade denaturing bis:acrylamide gels. Different gel densities were chosen depending on the size of the RNA. Di- and trinucleotide products were run on 33 % bis:acrylamide gels, 9 mer – 18 mer transcripts and primer were run on 23 % gels, and run-off transcripts were run on 15 % gels (made by dilution of 23 % gel mixture). After electrophoresis, the gels were incubated in a phosphoimaging screen cassette for 30 minutes or up to 2 days. The phosphoimage was visualised by Typhoon imager by GE Healthcare. Image analysis was performed using the software ImageQuant by GE Healthcare.

3 Chapter I: Bacterial and mitochondrial RNA Polymerases perform Non-canonical RNA capping

3.2 Introduction

With every new finding of molecular biology, many more questions arise. This is in no way different for the newly discovered non-canonical initiating nucleotide (NCIN) capping of bacterial and eukaryotic RNA. NCIN-capped RNA bears a nucleotide-containing small molecule, such as a cofactor or small metabolite, or a cell wall precursor, instead of a 5'-canonical NTP. Due to the uniqueness of the classic eukaryotic 5'-RNA-cap, RNA capping was believed to have originated from an evolutionary time point after divergence of pro- and eukaryotes. In 2009 NCIN caps were first discovered on bacterial transcripts, and later shown to also be present on eukaryotic and viral RNA. The conservation of NCIN-capping in different kingdoms of life implies an important function. First findings allow to make assumptions about a physiological role of bacterial NCIN-caps in regulation of growth and virulence. There is so far no evidence for preserved function of NCIN-caps throughout evolution.

A few years after the discovery of the existence of non-canonical caps we (and others) have elucidated the mechanism by which RNAP caps transcripts.

3.2.1 Chronology of the Discovery

The understanding of versatility and functionality of the RNA molecule in the cell has been extended considerably in the last decades (Chen *et al.*, 2009; Repoila and Darfeuille, 2009; Waters and Storz, 2009). RNA seems to fulfil a much higher variety in functions than the molecule's primary structure conveys. Thus, in 2009 the group around David Liu performed a screening of *E. coli* (Gram negative) and *Streptomyces venezuelae* (Gram positive) cellular RNA in an attempt to find previously unknown small molecule conjugates on RNA (Chen *et al.*, 2009; Kowtoniuk *et al.*, 2009). Their method was a combination of size exclusion chromatography, chemical and enzymatic processing of RNA and Mass Spectrometry and yielded besides many known RNA conjugates (such as amino acids on tRNA), and base modifications such as methylation, around 30 new small molecule conjugates for both organisms.

Liu's group focused on the determination of two of the unknown conjugates found in both species and identified dephospho-Coenzyme A (DP-CoA) and its succinyl- acetyl- and methylmalonyl-thioester derivatives, as well as nicotinamide adenine dinucleotide (NAD⁺).

NAD⁺-RNA was further determined to be present at high abundance in the bacterial cell, comparable in number to an individual tRNA. Using further analysis they were able to localise both conjugates to the 5'-end of RNA, thus the parallel to eukaryotic cap was drawn (Chen *et al.*, 2009; Kowtoniuk *et al.*, 2009; Cahová *et al.*, 2015).

First studies aimed at identifying which types of genes are frequently NADylated (NAD⁺-modified). Cahová *et al.* found that in *E. coli*, most frequently sRNAs with functions for regulation of metabolism and stress response are NADylated. Further, 5'-mRNA fragments (size < 200 nt) of genes involved in metabolic pathways as well as stress response were enriched by their method NAD captureSeq (Cahová *et al.*, 2015). With 13% of transcripts the highest NADylated RNA in *E. coli* was found to be RNA1, sRNA for replication control of plasmid ColE1. NAD captureSeq makes use of transfer reactions to enzymatically exchange the nicotinamide mononucleotide (NMN) moiety of NAD⁺ to a penynyl group from pentynol (transglycosylation) which can be chemically “clicked” to a biotin tag (copper-catalysed azide-alkyne cycloaddition) which was then captured on streptavidin beads. The captured RNA was reverse transcribed, PCR amplified, and the cDNA sequenced. The same method was tested on the yeast *Saccharomyces cerevisiae*, yielding the first proof of NADylated RNA in eukaryotic cells. mRNAs and mitochondrial RNAs were enriched, mainly of genes involved in translation and mitochondrial activity, such as cytochrome c oxidase subunit 2 and F0 ATP synthase subunit c (Walters *et al.*, 2017). Shortly after this, NAD captureSeq was applied on HEK293T (human kidney) cells revealing that mainly small nuclear (snRNA) and small nucleolar (snoRNA), as well as small Cajal body (scaRNA) RNAs are NADylated in mammals (Jiao *et al.*, 2017). The finding that this RNA modification exists in both prokaryotes and eukaryotes spiked the interest in the discovery in hope to find unknown, ancestral mechanisms and pathways. This alternative capping was suggested to implement an unknown mechanism of epigenetic regulation (Kiledjian, 2018). *In vitro* experiments suggested that NCIN capping can also be expected on human mitochondrial, as well as viral RNA (Huang, 2003; Julius, Riaz-Bradley and Yuzenkova, 2018), which was recently verified *in vivo* (Bird, Basu, Kuster, Ramachandran, Grudzien-Nogalska, Towheed, *et al.*, 2018; J. Wang *et al.*, 2019). For this, new methods were developed. Bird *et al.* extended the Boronate affinity electrophoresis method developed by Nübel *et al.* in 2016 with hybridisation to radioactive probes (Nübel, Sorgenfrei and Jäschke, 2017; Bird, *et al.*, 2018). Wang *et al.* invented CapQuant, a quantification method that relies on HPLC enrichment of non-canonical nucleotides and LC-MS/MS (J. Wang *et al.*, 2019). They also used cap-analysis gene expression (CAGE) for identification of TSS, though this method relies on classical

eukaryotic caps. At last, NADylated RNA was also identified in the plant *Arabidopsis* (Y. Wang *et al.*, 2019; Zhang *et al.*, 2019), and *Bacillus subtilis* as well as *Staphylococcus aureus* (Frindert *et al.*, 2018; Morales-Fillooy *et al.*, 2019), using NAD captureSeq and another new qualitative NAD⁺-RNA detection method, NAD tagSeq. Here, NAD⁺ is enzymatically replaced by a synthetic RNA tag, detected with a complementary DNA probe, and RNA is sequenced (Zhang *et al.*, 2019).

Besides discovery of capped RNAs, also potential dedicated cap removal (decapping) enzymes were identified. An *E. coli* protein of the family of NUDIX (nucleotide diphosphate bound to a moiety X) hydrolases, NudC, had previously been known as housekeeping enzyme for NAD(H) homeostasis (Frick and Bessman, 1995). NudC cleaves the disphosphate bond in NAD(H), producing nicotinamide mononucleotide (NMN) and AMP. Its' strong affinity to NADylated RNA was shown, which makes it a great tool to study NAD⁺-capping (Cahová *et al.*, 2015; Höfer *et al.*, 2016). Another NUDIX family protein was found to be involved in NAD⁺-cap removal and pyrophosphohydrolysis of triphosphorylated transcripts in *B. subtilis*, BsRppH (Frindert *et al.*, 2018). Interestingly, the *E. coli* homologue RppH is a pure pyrophosphohydrolase (Foley *et al.*, 2015), indicating a potential difference in NCIN-RNA degradation between Gram positive and Gram negative bacteria. NudC homologues were found in eukaryotic cells (Zhang *et al.*, 2016). These might constitute one pathway of cap removal in eukaryotes, while another protein family, DXO, was also shown to decap both classical caps and NAD⁺-cap, albeit by a different mechanism (Jiao *et al.*, 2017).

The mechanism of NCIN-capping was soon identified using biochemical assays (Bird *et al.*, 2016; Julius and Yuzenkova, 2017). The presented research in this thesis is part of these studies and will be discussed further in this chapter.

Table 3.1. shows a timeline of all studies conducted on NCIN-capping and decapping until the present time.

Table 3.1 Chronology of discoveries on NCIN-capping and decapping, including research papers but not reviews.

publication date	Authors	Finding	Methods
May-09	Kowtoniuk <i>et al.</i>	multiple small molecule conjugates on <i>E. coli</i> and <i>S. venezuelae</i> RNA, particularly CoA species	alkaline/nucleophile chemical breakdown of RNA, LC/MS
Dec-09	Chen <i>et al.</i>	NAD ⁺ -linked RNA on subpopulations of small (< 200 nt) transcripts cannot be installed by RNAP <i>in vitro</i>	nuclease P1 digestion, LC/MS
Mar-15	Cahová <i>et al.</i>	<i>E. coli</i> NAD ⁺ -modification mainly on sRNAs and mRNA fragments, quantification of these, NAD-linked transcripts are resistant to RppH and RNase E cleavage, but susceptible to NudC	NAD captureSeq, <i>in vitro</i> transcription, <i>in vitro</i> nuclease assays
Jul-16	Bird <i>et al.</i>	Non-canonical initiating nucleotides are substrates to <i>E. coli</i> and yeast RNAP, decapping by NudC with crystal structure, higher prevalence in stationary phase	<i>in vitro</i> transcription, RNA isolation, $\Delta nudC$ mutant
Sep-16	Höfer <i>et al.</i>	crystal structure of <i>E. coli</i> NudC with NAD ⁺ or NMN, NudC prefers NAD ⁺ -RNA over free NAD ⁺ , RNA binding is unspecific	Crystallography, mutation studies, <i>in vitro</i> NAD ⁺ -hydrolysis, next generation sequencing
Aug-16	Zhang <i>et al.</i>	Crystal structure of <i>E. coli</i> NudC with NAD ⁺ , NAD ⁺ hydrolysis is metal ion-dependent, eukaryotic NudC-homologues decap RNA <i>in vitro</i>	Crystallography, mutation studies, <i>in vitro</i> decapping
Sep-16	Nübel, Sorgenfrei and Jäschke	Boronate-affinity electrophoresis as new method to analyse and isolated NAD ⁺ -RNA	acryloylaminophenyl boronic acid- polyacrylamide gels to retain NADylated RNAs, gel elution
Jan-17	Walter <i>et al.</i>	NAD ⁺ -cap on yeast nuclear and mitochondrial pre-mRNA	NAD captureSeq (NAD ⁺ on RNA is enzymatically replaced with biotin, streptavidin-captured RNAs are reversed transcribed, cDNA is sequenced)

Mar-17	Jiao <i>et al.</i>	In mammals, predominantly NAD ⁺ -snRNA and NAD ⁺ -snoRNA, DXO proteins as new class of NCIN-removing enzymes, crystal structure of DXO with 3'-pNAD, NADylation interferes with translation	NAD captureSeq, <i>in vitro</i> transcription, <i>in vitro</i> decapping, <i>in vivo</i> reporter-mRNA translation (luciferase) and fluorescence detection, DXO-KO
May-17	Julius and Yuzenkova	bacterial RNAP incorporates various cofactors of metabolism (NAD(H), NADP, DP-CoA, FAD) and cell wall precursors (UDP-Glc, UDP-GlcNAc) <i>in vitro</i> , NCIN-capping supports promoter escape, protein structural determinants of NCIN-capping (σ region 3.2, β Rif-region)	<i>in vitro</i> transcription, mutation studies
Apr-18	Julius, Riaz-Bradley and Yuzenkova	mitochondrial RNAP incorporates metabolites <i>in vitro</i>	<i>in vitro</i> transcription
May-18	Vvedenskaya <i>et al.</i>	CapZyme-Seq as new method for NAD ⁺ -RNA detection and quantitation, promoter structural determinants for NAD ⁺ -capping	<i>in vitro</i> transcription of promoter library, decapping enzyme-and barcode DNA-adapter-based quantitation of transcript, reverse transcription, sequencing
Jul-18	Bird <i>et al.</i>	NAD(H)-capping by mitochondrial RNAP is more efficient than bacterial and yeast RNAP and reflect intracellular NAD(H) levels	<i>in vitro</i> transcription, nuclease cleavage, boronate affinity electrophoresis, hybridisation to radioactive probes
Aug-18	Frindert <i>et al.</i>	Detection of NAD ⁺ -RNA in <i>B. subtilis</i> (Bs) are mainly mRNAs, BsRppH acts as decapping enzyme <i>in vitro</i> and <i>in vivo</i> , protein and promoter structural determinants	nuclease digest, decapping, LC/MS, NAD captureSeq, <i>in vitro</i> transcription, mutation studies
Oct-18	Grudzien-Nogalska <i>et al.</i>	NAD-capQ as new method to detect and quantitate NAD ⁺ -RNA, level of NAD ⁺ -RNA is proportional to cellular NAD ⁺ -level in human cells	nuclease P1 digestion, colorimetric detection of NAD ⁺ , <i>in vitro</i> transcription and decapping, Δ nudC, Δ rai1, Δ DXO1, DXO-KO

Jun-19	Wang <i>et al.</i>	Identification of NAD ⁺ -capped mRNA of nuclear, mitochondrial, but not chloroplast genes of <i>Arabidopsis</i>	NAD captureSeq
Jun-19	Zhang <i>et al.</i>	New method NAD tagSeq to identify genes that yield NADylated RNA	NAD tagSeq (NAD ⁺ is enzymatically replaced by synthetic RNA tag and enriched by complementary DNA probe, bound RNA sequenced)
Jul-19	Wang <i>et al.</i>	<i>in vivo</i> evidence of FAD, UDP-Glc and UDP-GlcNAc caps on bacterial, yeast, mammalian and viral RNA	CapQuant (nuclease P1, HPLC enrichment of cap nucleotides, LC-MS/MS for quantification), cap-analysis gene expression (CAGE) for TSS identification (pull-down, reverse transcription, and bioinformatic analysis), RT-qPCR
Sep-19	Morales-Filloy <i>et al.</i>	NADylation of antitoxin RNAlII in <i>S. aureus</i> modulates virulence	NAD captureSeq, RNA Seq

3.2.2 Mitochondrial transcription is dissimilar to bacterial and eukaryotic transcription

After NAD⁺- and DP-CoA-capped bacterial transcripts were discovered, similar methods were used to detect NADylated RNA in yeast, revealing that yeast NADylated RNA were nuclear pre-mRNAs (unprocessed transcripts) and unprocessed mitochondrial transcripts. The yeast mitochondrial genome (mtDNA) is circular and encodes 30-40 genes, including ribosomal RNAs, 24 tRNAs, and genes for enzymes of oxidative phosphorylation (ATP synthase subunits, cytochrome c oxidase subunits (COX), apocytochrome b, and in some yeast NADH-ubiquinone oxidoreductase subunits, ribosomal protein, and RNA subunit of mitochondrial RNase P). *S. cerevisiae* carries 50-200 mitochondrial chromosomes, depending on growth phase. mtDNA is about 85 kb large (Freel, Friedrich and Schacherer, 2015). Yeast mtDNA bears 28 promoters, all of which are +1A. Interestingly, mutation of +1 position was shown not to affect transcription. Therefore, authors hypothesised that “ATP sensing” is a means of transcription regulation by the availability of ATP in the mitochondrion. Furthermore, K_M of POLMRT to initiating ATP varies on different promoters, making them more or less sensitive to regulation by ATP concentration (Deshpande and Patel, 2014). Transcripts of 21S rRNA gene and *cox2*, which were identified as highly NADylated *in vivo* (Walters *et al.*, 2017), are reported to show lower sensitivity to regulation by ATP (Deshpande and Patel, 2014).

The mammalian mitochondrial genome is circular, 16.6 kb, and encodes 13 proteins, 22 tRNAs and 2 rRNAs. Uneven distribution of purines/pyrimidines causes one strand to be heavier (H strand) than the other (L strand). Transcription takes place starting at one promoter per strand, the heavy strand- or light strand promoter (hsp or lsp), yielding long polycistronic transcripts which require processing. Both promoters are located at a non-coding region together with origin of replication of H strand (Taanman, 1999). Both promoters are +1A (Morozov and Temiakov, 2016). Most genes are located on H-strand, starting with two rRNA genes, 14 tRNAs and 12 protein coding genes. The lsp-transcript starts with 8 tRNAs, and carries one protein gene. All mtDNA encoded proteins belong to the oxidative phosphorylation pathway (Taanman, 1999). Replication also starts from transcripts on lsp (Chang and Clayton, 1985).

Bacterial and eukaryotic cellular RNAPs are structurally similar multisubunit enzymes. Mitochondrial RNAP (POLMRT) however, is a single subunit enzyme with structural relation closer to bacteriophage transcriptases such as T7 RNAP. The C-terminal domain of the single subunit RNAPs forms a form a “right-hand” structure with thumb, palm and finger domains

wrapped around DNA. The catalytic centre is also located in this domain. Polymerisation requires two divalent cations. An α -helix of the finger domain is involved in promoter melting and remains at the downstream side of the transcription bubble. Promoter contact of POLMRT is via N-terminal domain and comprises an AT-rich recognition loop and an intercalating hairpin which is also involved in promoter melting. The specificity loop in the C-terminal domain stabilises the open complex and localises the template strand to the catalytic centre (Hillen, Temiakov and Cramer, 2018).

Mitochondrial transcription initiation requires two essential transcription factors, mitochondrial transcription factor A (TFAM) and mitochondrial transcription factor B2 (TFB2M). As shown in figure 3.1, transcription is initiated by TFAM binding promoter DNA -10 to -15 bp upstream of TSS. It induces a U-turn in the promoter and recruits POLMRT which binds on its N-terminus. POLMRT contacts DNA on a region at approximately -50 to -60 relative to TSS. TFB2M binds POLMRT and induces structural isomerisation to open complex and promoter melting (D'Souza and Minczuk, 2018; Barchiesi and Vascotto, 2019).

In mitochondrial transcription elongation, initiation factors are lost and mitochondrial transcription elongation factor (TEFM) binds POLMRT, which enhances transcription processivity and enables transcription of promoter distal genes. Especially transcription of RNAs high in secondary structure (such as tRNA) requires support by TEFM (Barchiesi and Vascotto, 2019). TEFM forms a “sliding clamp” structure around DNA and interacts with POLMRT via its C-terminus (D'Souza and Minczuk, 2018). 39S ribosomal protein L12 (MRPL12) supports processivity of elongation (Barchiesi and Vascotto, 2019).

In absence of TEFM transcription from *lsp* can be prematurely terminated at a site called conserved sequence block 2 (CSB2), a non-coding region. The transcript that results from this, premature termination can serve as primer for DNA replication (Chang and Clayton, 1985; D'Souza and Minczuk, 2018).

Mitochondrial transcription termination is not fully understood. Likely, mitochondrial termination factor 1 (MTERF1) binds mtDNA between hsp and tRNA^{Leu}, inducing unwinding of DNA and base flipping. The elongation complex clashes with this termination structure (D'Souza and Minczuk, 2018). Other studies, however, have shown that MTERF1 is expendable for mitochondrial transcription termination (Barchiesi and Vascotto, 2019). Another potential termination mechanism depends on folding of the RNA-DNA hybrid into a structure that halts the elongation complex. This folding occurs at a G-rich sequence,

neighbouring by a short spacer and an AU-rich stretch (Wanrooij *et al.*, 2012; Mukundan and Phan, 2013).

Long polycistronic mitochondrial transcripts need to be processed. In the “tRNA punctuation model” of mtRNA processing, a first step of processing is cleavage of 5’- and 3’-ends of tRNAs by RNase P and Z, respectively (D’Souza and Minczuk, 2018). RNase P is predominantly described to be involved in tRNA maturation. It is a multiprotein complex that is encoded in the nucleus but acts on mitochondrial RNA. One subunit of RNase P, MRPP2, has an NAD⁺-binding site (Lopez Sanchez *et al.*, 2011).

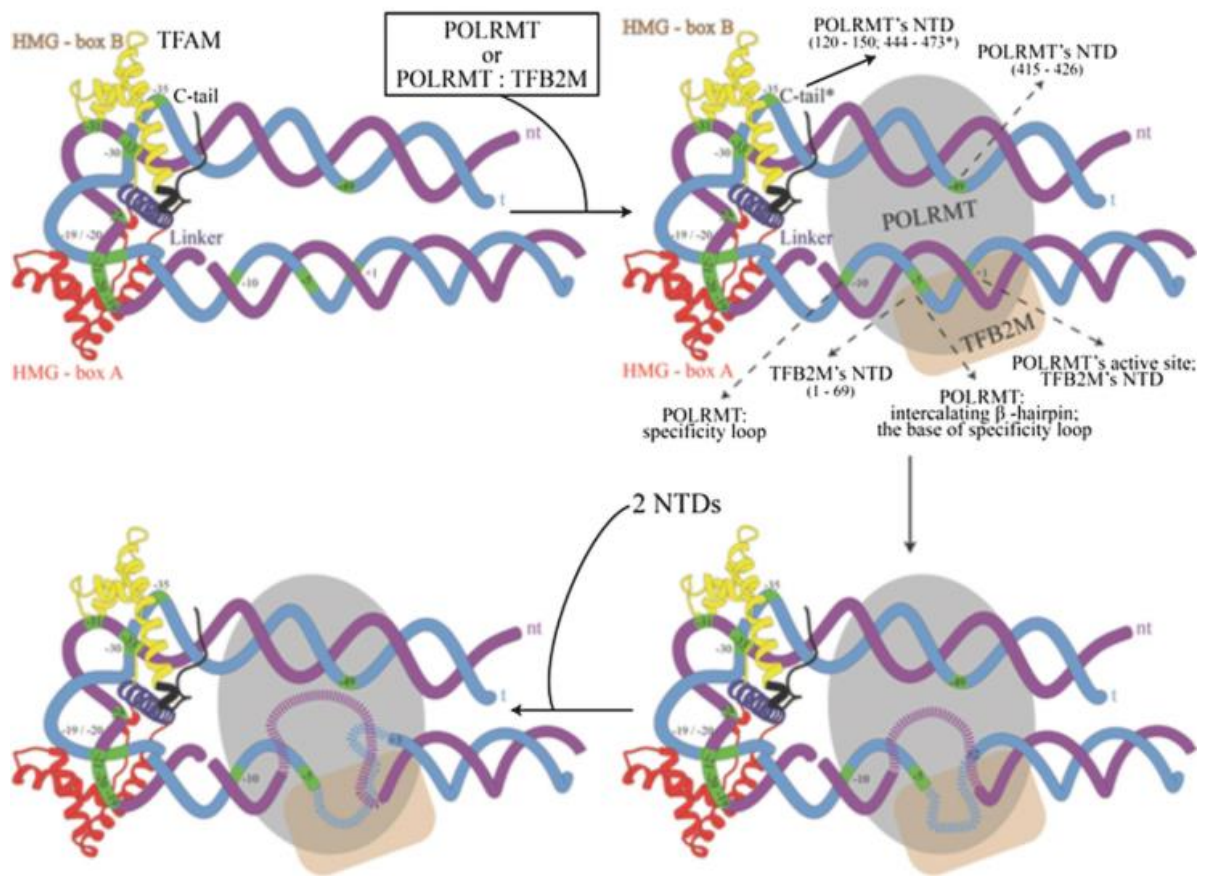


Figure 3.1 From (Drakulic, Cuellar and Sousa, 2018). Mitochondrial transcription initiation. Transcription requires binding of TFAM to promoter DNA. POLRMT recognises the characteristic U-turn of the TFAM-promoter complex. TFB2M binding to POLRMT causes promoter melting and transcription can commence.

Additionally, several cellular Fas-activated serine/threonine kinases (FASTKs) might be involved in maturation of mitochondrial transcripts. Deletion of these kinases was shown to lead to accumulation of precursor mtRNAs, indicating a role in processing (D’Souza and Minczuk, 2018). After cleavage from the primary transcript, mRNAs are polyadenylated by polyadenylic acid RNA polymerase (mtPAP). Polyadenylation is assisted by an RNA chaperone, leucin-rich penticopeptide-rich domain containing protein (LRPPRC), which

interacts with stem-loop interacting RNA-binding protein (SLIRP) to enhance stability of both LRPPRC-mRNAs (D'Souza and Minczuk, 2018). In case of some mRNAs polyadenylation completes an otherwise absent stop codon. Besides that, the effect of polyadenylation in mammalian mitochondria is RNA-specific and cannot be generalised to lead to degradation, as in bacteria, or stability, as in the eukaryotic cytosol (D'Souza and Minczuk, 2018).

3.3 Aims

We hypothesise that the small-molecule conjugates found on 5'-RNA are utilised by bacterial RNAP to initiate transcription, and are thus incorporated at the 5'-end as a non-canonical initiating nucleotide. Our results should further shed light on several aspects of the capping mechanism, such as promoter and protein structural elements. We aim to verify that RNA-capping with nucleotide-containing molecules is an ability of various DNA-dependent RNA Polymerases, specifically bacterial RNAP, Bacterial primase, and mitochondrial POLMRT.

3.4 Results

3.4.1 *Bacterial RNA polymerase incorporates adenine containing cofactors for transcription on +1A promoters*

To test the hypothesis that the nucleotide-containing molecules at RNA 5'-end found by Liu have been attached via transcription initiation, we performed *in vitro* transcription assays. This method was inspired by early research on RNAP mechanism many years ago, where coenzymes NAD⁺, FAD, and DP-CoA were incorporated into RNA by T7 RNAP (Malygin and Shemyakin, 1979). These metabolites bear a free 3'-OH group from the ribose of adenosine, and α - and β -phosphates, but in place of γ -phosphate, another moiety X (such as nicotinamide or riboflavin) is attached. Here, we assess whether RNAP incorporates these non-canonical substrates in a template-dependent fashion. We hypothesised that the adenine-containing metabolites could substitute for ATP on a +1 A promoter. This requires base pairing between non-template strand dT and incoming A (or ADP-part of metabolites). In a first experiment with metabolites and α (P³²)-UTP on bacteriophage promoter *t7al* dinucleotide abortive products were synthesised by EcRNAP with all tested nucleotide-containing metabolites except NADP (see figure 3.2). The gel clearly shows the variation of migration of products on denaturing bis:acrylamide gel, caused by size and charge differences of the X moieties of initiator molecules. A further observation peaked our interest: Variations

in radioactive signal intensity indicates that the efficiency of dinucleotide formation differs, which means that RNAP has varying affinities to the substrates.

We tested this hypothesis with kinetics experiments (figure 3.3). Michaelis-Menten constant K_M was determined for substrates for ATP ($90 \pm 11 \mu\text{M}$), NAD^+ ($358 \pm 67 \mu\text{M}$), and NADH ($380 \pm 72 \mu\text{M}$). Table 3.2 compares K_M values with cellular concentration of these metabolites in *E. coli* cells growing in complex medium with glucose as carbon source, measured from log-phase cells (Bennett, Kimball and Gao, 2009). If K_M of RNAP to a substrate is lower than the cellular concentration, this is an argument in favour of this process taking place *in vivo*. For both ATP and NAD^+ , cellular concentration is higher than K_M . NADH is present at lower concentrations *in vivo*. However, the ratio of NAD^+ to NADH is highly variable, dependent on many factors (such as oxygen levels) and underlies inherent measurement difficulty. Other studies challenge numbers presented by Bennet *et al.* and describe $\text{NAD}^+:\text{NADH}$ ratio is often equal (Sun *et al.*, 2012). Total NAD(H) content was determined to range between a minimal 0.039 mM and maximal 8.49 mM in *E. coli* over various growth stages (Zhou *et al.*, 2011).

Since RNA1 was shown to be most frequently capped *in vivo* (Cahová *et al.*, 2015), we tested NCIN-capping on the *rnaI* promoter (figure 3.3). This promoter was then modified to allow for production of a transcript of specific size (*rnaI* mod, figure 3.3 A). Interestingly, while NADP could not initiate transcription on *t7aI* (figure 3.2 B), it was incorporated in both abortive as well as longer RNA products from *rnaI* and *rnaI* mod (figure 3.3 C). This led us to suspect that promoters might contain signals to regulate NCIN-capping.

Table 3.2 Comparison of cellular concentration and K_M of metabolites used as initiating substrates for transcription. Metabolite concentrations were tested by Bennet *et al.* in exponentially growing *E. coli* cells on complex medium, with glucose as carbon source (Bennett, Kimball and Gao, 2009). Michaelis-Menten constant was tested by *in vitro* transcription on *rnaI* promoter giving excess initiator (0.5 mM) and $\alpha(\text{P}^{32})\text{-CTP}$ to *E. coli* RNAP. Transcription efficiency was determined via phosphor-imaging using ImageQuant. (Julius and Yuzenkova, 2017)

substrate	ATP	NAD^+	NADH
Cellular concentration (mM)	9.6	2.6	0.08
K_M in transcription (mM)	0.09	0.36	0.38

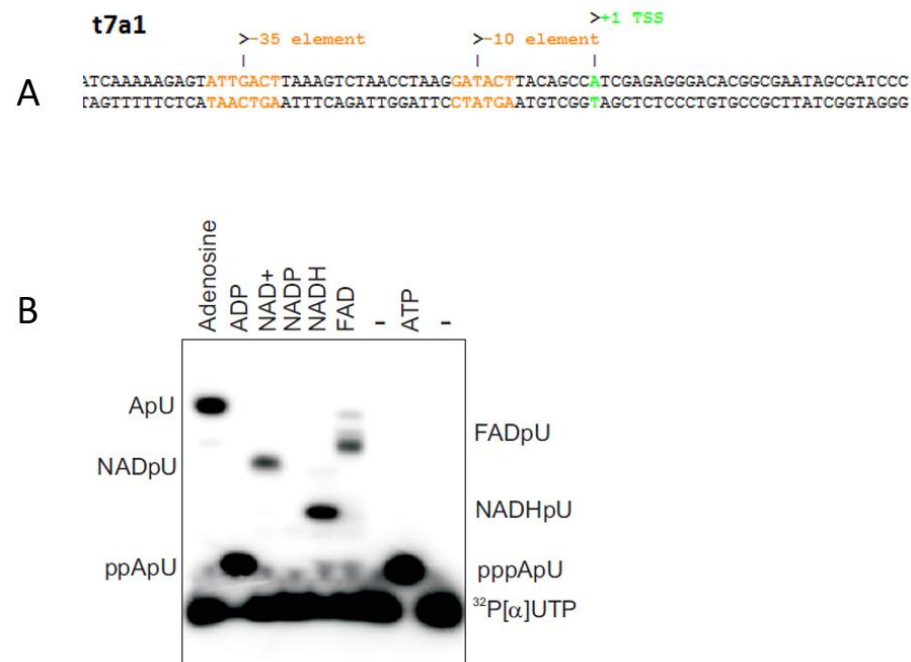


Figure 3.2 Promoter-dependent transcription initiation by EcRNAP using different adenine-containing metabolites. A sequence of *t7a1* promoter with -35 and -10 promoter elements (orange) and TSS (green). B Dinucleotide products of adenine-containing initiator, given at high concentration (0.5 mM), and radiolabelled $\alpha(\text{P}^{32})$ -UTP as second nucleotide substrate. Results were obtained in collaboration with Dr Yulia Yuzenkova (Julius and Yuzenkova, 2017).

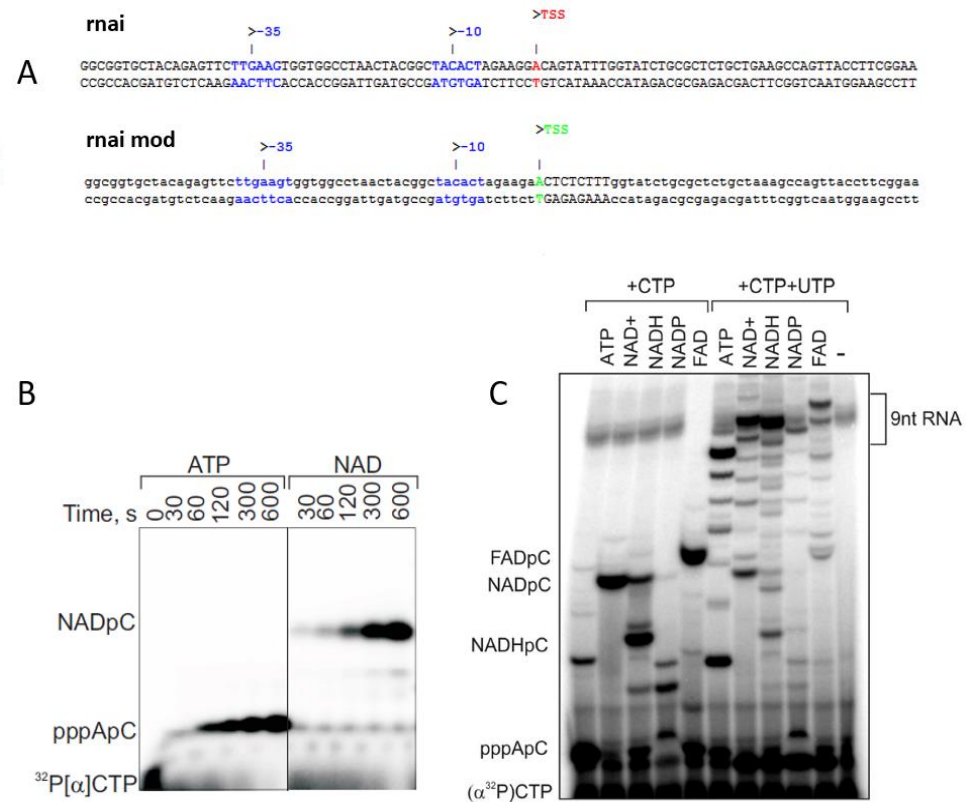


Figure 3.3 Metabolite-capping on promoter *rnai*. A Sequences of promoters *rnai* and modified version *rnai mod*, which allows for transcription of precise 9-nt product when giving adenosine-containing initiator (+1A), $\alpha(\text{P}^{32})$ -CTP (+2C) and UTP. B Denaturing polyacrylamide gel of time-dependent synthesis of dinucleotide product initiated by ATP or NAD^+ . C Production of dinucleotide and 9-nt products initiated by cofactors on *rnai mod*. Results were obtained in collaboration with Dr Yulia Yuzenkova (Julius and Yuzenkova, 2017).

3.4.2 Uridine-containing cell wall precursors initiate transcription on +1U promoters

Encouraged by the results for adenine-containing metabolites, we aimed to investigate whether another TSS would allow for capping with other nucleotide containing small molecules. Among the most abundant in growing bacterial cells is uridine diphosphate N-acetyl glucosamine (UDP-GlcNAc), an intermediate of cell wall synthesis (9.2 mM (Bennett, Kimball and Gao, 2009)). Promoter *acnA* (of *E. coli* aconitate hydratase A gene) is +1U promoter (Cunningham, Gruer and Guest, 1997). We tested incorporation of UDP-GlcNAc, uridine diphosphate glucose (UDP-Glc) and uridine diphosphate N-acetyl-muramic acid-pentapeptide (UDP-MurNAc-AA5). UDP-MurNAc-AA5 is a larger cell wall precursor which consists of five amino acids covalently bound UDP-MurNAc. Further, we included smaller uridine-analogues UDP and UMP (figure 3.4). Similarly to adenosine-containing coenzymes, uridine-containing cell wall precursors can be incorporated into RNA, both in abortive transcription as well as longer transcripts. The large UDP-MurNAc-AA5 was not able to initiate transcription on *acnA* (figure 3.4 B). Table 3.3 compares K_M of cell wall precursors as substrates for RNAP with their cellular concentration, indicating that the reaction potentially takes place *in vivo*.

Table 3.3 Comparison of cellular concentration and K_M of UTP and cell wall precursors used as initiating substrates for transcription. Metabolite concentrations were tested by Bennet *et al.* in exponentially growing *E. coli* cells on complex medium, with glucose as carbon source (Bennett, Kimball and Gao, 2009). Michaelis-Menten constant was tested in promoter-dependent *in vitro* transcription giving excess initiator (0.5 mM) and $\alpha(P^{32})$ -CTP to *E. coli* RNAP. Transcription efficiency was determined via phosphor-imaging using ImageQuant. (Julius and Yuzenkova, 2017)

substrate	UTP	UDP-Glc	UDP-GlcNAc
Cellular concentration (mM)	8.3	2.5	9.2
K_M in transcription (mM)	0.12	0.33	0.30

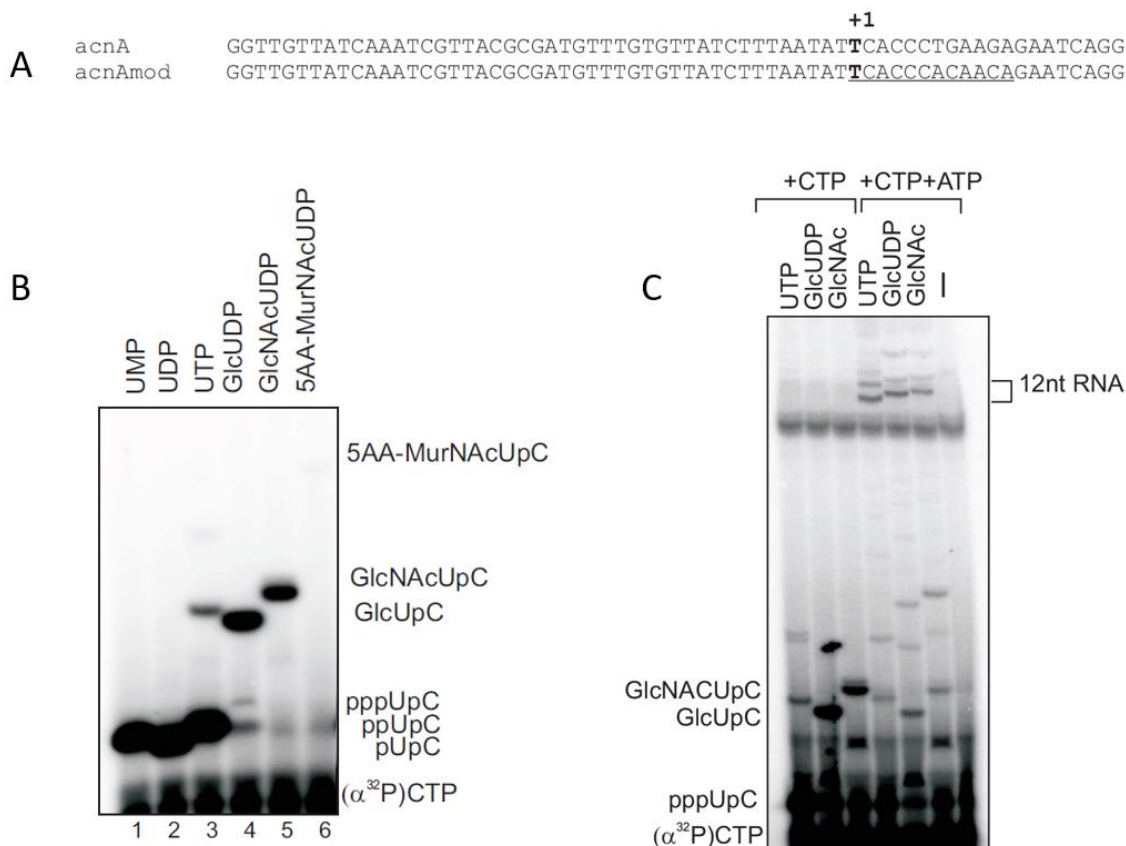


Figure 3.4 Transcription initiation by uridine-containing molecules on promoter *acnA*. A. sequence of wildtype *acnA* promoter and modified version, *acnA mod*, which allows for transcription of 12 nt products. For this uridine-containing initiator, radiolabelled $\alpha(P^{32})$ -CTP (+2C) and ATP are added. B. Resolution of dinucleotide products initiated by different uridine-containing molecules on denaturing bis:acrylamide gel (33%). C. Resolution of dinucleotide (left) and 12mer products (right), initiated by UTP or cell wall precursors on a 23% bis:acrylamide gel. Results were obtained in collaboration with Dr Yulia Yuzenkova (Julius and Yuzenkova, 2017).

3.4.3 Promoter -1 position does not affect NADylation efficiency

Cofactors such as NAD^+ and FAD are essentially dinucleotides. We had to consider the possibility that the non-nucleotide moiety of the cofactor might undergo base pairing or other specific non-covalent interaction with promoter position -1. Therefore, we tested whether mutation of -1 on a given template would affect the efficiency of incorporation of cofactor NAD^+ . As can be observed in figure 3.5, mutation of position -1 on the promoter *rnaI mod* from its wildtype -1A to -1C, -1 T, or -1 G affected incorporation of both, ATP and NAD^+ , similarly. This argues, that -1 template position affects transcription in general, but not NADylation specifically. Other studies have contradicted these results as discussed later.

A

RNAImod	GTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACTCTCTTTGGTATCTGCGCTCTGC
RNAImod -1C	GTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGCACTCTCTTTGGTATCTGCGCTCTGC
RNAImod -1T	GTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGTAACTCTCTTTGGTATCTGCGCTCTGC
RNAImod -1G	GTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGAACTCTCTTTGGTATCTGCGCTCTGC

B

template	K _m [ATP], μM	K _m [NAD ⁺], μM
-1A	90±11	358±67
-1C	27±8	85±18
-1T	30±5	122±27
-1G	127±18	357±46

Figure 3.5 Effect of template position -1 on incorporation of canonical substrate ATP vs non-canonical substrate NAD⁺. A. Promoter *mai* mod was modified at -1 position relative to TSS. B. Analysis of enzyme kinetics for both initiating substrates on the different promoter variants shows that a pyrimidine (C or T) at -1 enhances incorporation of both substrates in comparison to wildtype -1A promoter. Exchange of -1A to -1G does not change K_M of NAD⁺ significantly. Values are mean ± standard deviation of K_M values of three independent experiments. (Julius and Yuzenkova, 2017)

3.4.4 σ factor might the limit size of the NCIN-cap

We have seen that some nucleotide-containing molecules are more efficient in transcription initiation than others. Interestingly, size and incorporation efficiency seem to have a proportional relationship, with bulkier molecules being less efficiently incorporated than smaller ones. σ subunit of RNAP binds at promoter sites close to TSS. It is known that σ^{70} region 3.2 is involved in transcription initiation (Pupov *et al.*, 2014). In RNAP crystal structure in complex with DNA and NAD⁺pC (Bird *et al.*, 2016), σ^{70} region 3.2 is in close promoxity to the catalytic centre and to bound NAD⁺. This region is not contained in all σ factors and it is not conserved across the eubacteria (M. Paget, 2015). Therefore we tested which effect deletion of region 3.2 of σ^{70} will have on NCIN-capping *in vitro*. Figure 3.6 B and C shows that transcription efficiency by the mutant holoenzyme is slightly affected, however, incorporation efficiency of NCINs is unaltered as compared to wildtype holoenzyme. This finding was true for all tested NCINs, except for UDP-MurNAc-AA5. In figure 3.6 C we show that deletion of 3.2 facilitates incorporation of UDP-MurNAc-AA5 into the transcript. This might indicate that indeed, the 3.2 region sterically selects initiating substrates due to size or other unidentified biochemical interaction.

Bacterial σ are typically classified as σ^{70} -like or σ^{54} -like (M. Paget, 2015; Zhang and Buck, 2015). σ^S (or σ^{38}), the stationary phase σ , belongs to the σ^{70} -like factors. It contains a region 3.2, and it has similar promoter requirements to σ^{70} , but its' association to RNAP leads to expression of stationary phase-associated genes (Colland *et al.*, 2002). Interestingly,

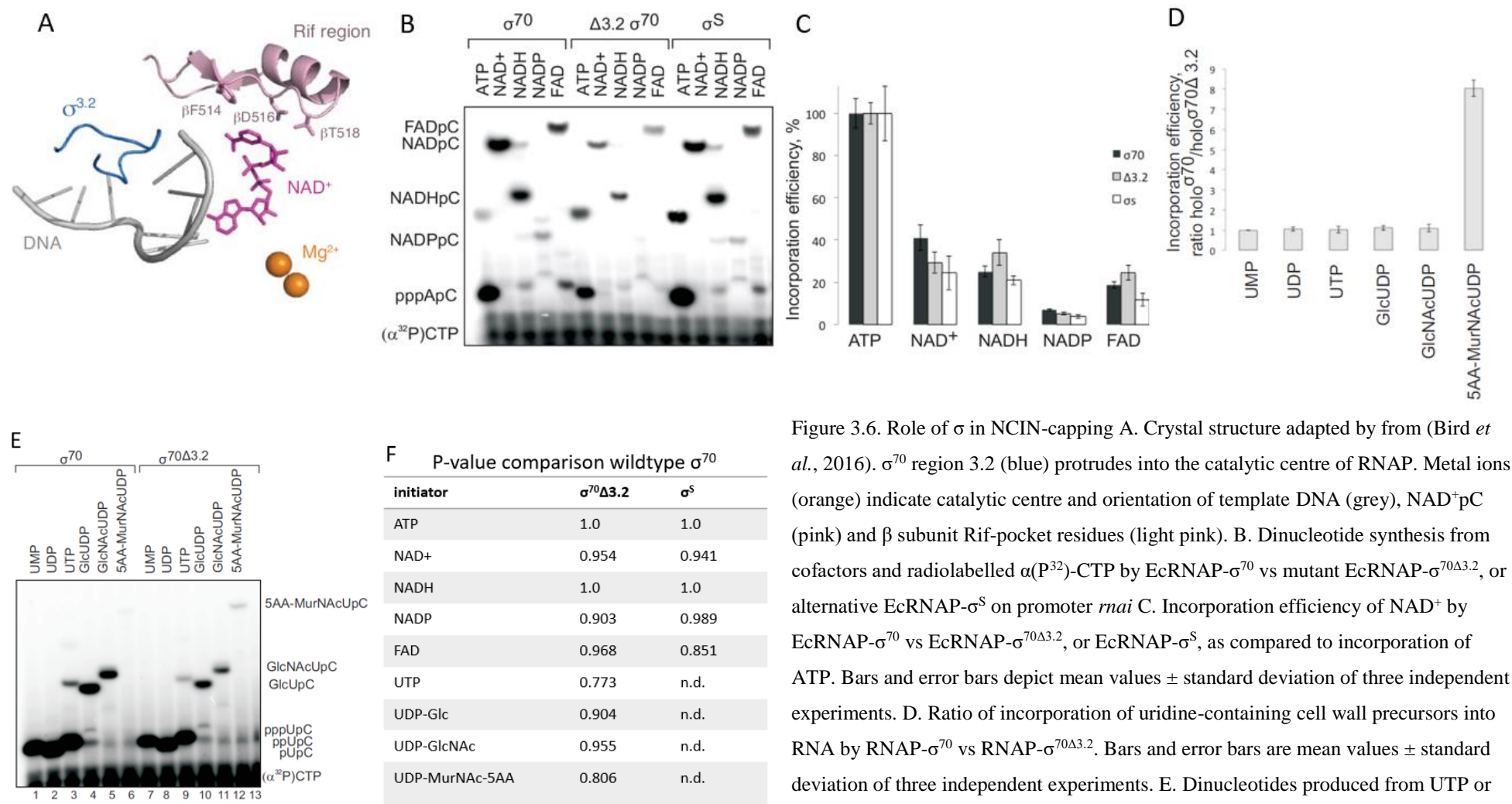


Figure 3.6. Role of σ in NCIN-capping A. Crystal structure adapted by from (Bird *et al.*, 2016). σ^{70} region 3.2 (blue) protrudes into the catalytic centre of RNAP. Metal ions (orange) indicate catalytic centre and orientation of template DNA (grey), NAD^+pC (pink) and β subunit Rif-pocket residues (light pink). B. Dinucleotide synthesis from cofactors and radiolabelled $\alpha(\text{P}^{32})\text{-CTP}$ by EcRNAP- σ^{70} vs mutant EcRNAP- $\sigma^{70\Delta 3.2}$, or alternative EcRNAP- σ^S on promoter *rnaI* C. Incorporation efficiency of NAD^+ by EcRNAP- σ^{70} vs EcRNAP- $\sigma^{70\Delta 3.2}$, or EcRNAP- σ^S , as compared to incorporation of ATP. Bars and error bars depict mean values \pm standard deviation of three independent experiments. D. Ratio of incorporation of uridine-containing cell wall precursors into RNA by RNAP- σ^{70} vs RNAP- $\sigma^{70\Delta 3.2}$. Bars and error bars are mean values \pm standard deviation of three independent experiments. E. Dinucleotides produced from UTP or uridine-containing cell wall precursors and $\alpha(\text{P}^{32})\text{-CTP}$ in absence or presence of $\sigma^{70.3.2}$. Here, incorporation of UDP-MurNAc-AA4 is increased in the mutant holoenzyme. F. P-values from experiments shown in C and D, comparing initiation efficiency of different substrates by EcRNAP- $\sigma^{70\Delta 3.2}$ or EcRNAP- σ^S to RNAP- σ^{70} . n.d.= not determined (Julius and Yuzenkova, 2017; Julius, Riaz-Bradley and Yuzenkova, 2018).

NADylation was found to be increased in stationary phase as compared to log phase cells in *E. coli* (Bird *et al.*, 2016). Therefore, we tested whether σ^S could be responsible for this observation. We initiated *in vitro* transcription with different cofactor caps using RNAP- σ^S holoenzyme, but we did not see a specific upregulation of NCIN-capping (figure 3.6 B and C).

3.4.5 *Rif-pocket residues might interact with NCIN-cap during initiation*

As shown in crystal structure (see figure 3.6 A) the rifampicin-binding pocket of RNAP (Rif-pocket), a site within the RNA-exit channel of EcRNAP where the antibiotic rifampicin binds, is close to the catalytic centre of the holoenzyme. Therefore we tested, if a) rifampicin addition to the assay affects NCIN-capping and b) if rifampicin-resistant mutants (RNAPs which have point mutations in the Rif-pocket) have an altered capping spectrum as compared to the wildtype (wt) RNAP. The mechanism of action of rifampicin is to block the RNA exit channel and thus to inhibit not initiation, but extension of short RNA chains. *In vitro*, we can observe an increased production of canonical abortive products and inhibition of longer products, when subinhibitory concentrations of rifampicin are added to transcription assays. We used *rnaI* mod template to see whether low concentration of rifampicin (5 $\mu\text{g/ml}$) would have the same effect on transcription initiation by NAD^+ . Figure 3.7 shows that, as expected, ATP-initiated 9mer transcripts are reduced, and the amount of abortive product is increased in presence of rifampicin. However, abortive products initiated by NAD^+ were not increased by rifampicin addition. NAD^+ -initiated 9mer transcription was reduced in presence of rifampicin, thus the antibiotic acts still as transcription inhibitor. The fact that it does not enhance the release of NADylated abortive products indicates that NAD^+ -initiation counteracts early abortive release, and thus promotes RNAP promoter escape.

We then tested rifampicin-resistant mutants of RNAP with single amino acid substitutions in the Rif-pocket (purified from the ASKA strain collection). Figure 3.8 shows that RNAP mutants produced relatively less NAD^+ -RNA as compared to wt RNAP, if relative incorporation efficiency of ATP to NAD^+ is calculated. However, residues Q513 and D516 seem to have the strongest detrimental effect on NAD^+ -capping. As can be seen in the crystal structure (figure 3.7 B), those residues are closest to NAD^+ .

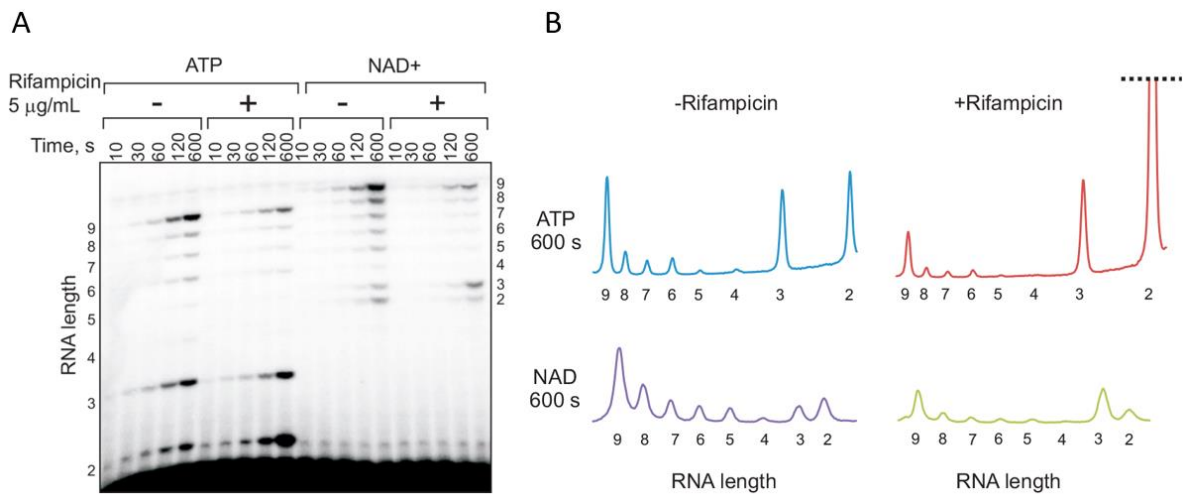


Figure 3.7 Initiation with NAD⁺ promotes promoter escape. Subinhibitory concentration of the antibiotic rifampicin leads to increased production of 2- and 3-mer abortive products when initiated with ATP, but not when transcription is initiated with NAD⁺. A. Bis:acrylamide gel electrophoresis of RNAP kinetics of 9-mer transcript from *rnaI* mod initiated by ATP (left) or NAD⁺ (right). Addition of rifampicin (5 μ g/mL, this is 10 times lower than inhibitory concentration), leads to increased production of short abortive product pppApC, but not NAD⁺pC. B. Depiction of the same experiment as signal traces across 60 sec bands. Results were obtained in collaboration with Dr Yulia Yuzenkova (Julius and Yuzenkova, 2017).

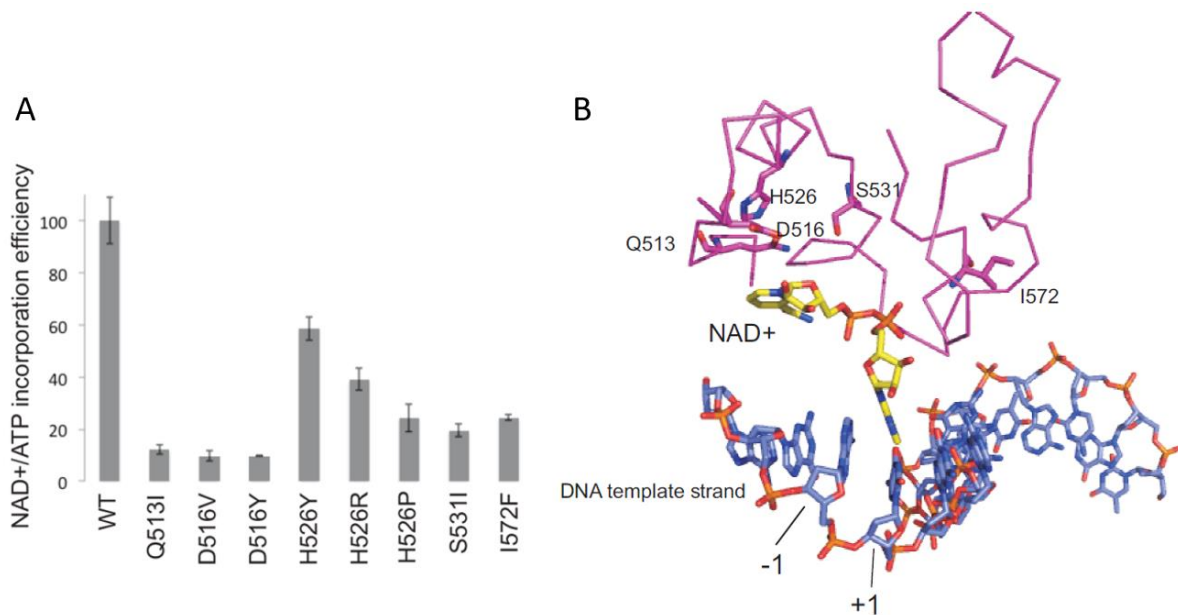


Figure 3.8 Rif-pocket residues are involved in NCIN-capping. A. Rif-resistant mutants of *E. coli* RNAP were tested for incorporation of NAD⁺ vs. ATP. Efficiency of incorporation of NAD⁺ by wildtype EcRNAP was normalised to 100%, all Rif-pocket mutants showed reduced NAD⁺-capping efficiency by at least 50%. Mutants of Q513 and D516 were found to be most detrimental to NAD⁺-incorporation, with around 90% loss of efficiency. P-values comparing relative transcription initiation efficiency with NAD⁺ to ATP of mutants against wildtype were $P < 0.001$ (Tukey algorithm). Bars and error bars are mean values \pm standard deviation of three independent experiments, Crystal structure of *T. thermophilus* RNAP in complex with DNA and NADpC. From (Bird *et al.*, 2016; Julius and Yuzenkova, 2017).

3.4.6 *NUDIX hydrolase NudC acts as decapping enzyme in vitro and can be used as a tool to verify NCIN-cap*

NudC has been established as NAD⁺-decapping enzyme in various studies (Cahová *et al.*, 2015; Bird *et al.*, 2016; Höfer *et al.*, 2016; Zhang *et al.*, 2016). It cleaves nicotinamide mononucleotide (NMN) from NAD⁺ by hydrolysis of phosphodiester bond. The product is AMP or monophosphorylated 5'-RNA starting with adenosine. We tested if NudC would also be applicable for decapping of other NCINs. This was also done in other studies with DP-CoA (Bird *et al.*, 2016). Here we show that NudC treatment results in RNA products of faster migration on denaturing gels if RNA was produced from initiation with NAD⁺ or FAD (figure 3.9 A). However, FAD seems to make a poor substrate, as we observe incomplete decapping. It is possible to verify cap structure also using nucleases RppH and terminator (commercially available). Terminator is a 5'-3'-exonuclease which is specific to 5'-monophosphorylated RNA. RppH cleaves pyrophosphate from triphosphorylated RNA substrates, creating a monophosphorylated end (Cahová *et al.*, 2015). This is susceptible to terminator treatment. Likewise, NudC creates monophosphorylated RNA from a capped transcript. We tested if transcripts are susceptible to terminator-mediated degradation in presence of NudC, and if they are susceptible to RppH treatment (figure 3.9 B). Visibly, ATP-initiated transcripts are not susceptible to NudC treatment, and capped transcripts are not susceptible to RppH treatment. Interestingly, the de-NADylated transcript was not susceptible to terminator (lane 9), while the de-FADylated transcript was (lane 14).

Potentially, terminator cannot access RNA in (transient) complex with RNAP and NudC, and incomplete decapping might be due to presence of RNAP. It was shown previously (Zhilina *et al.*, 2012), that 20mer transcripts are insusceptible to GreB-induced RNA cleavage, while 21-24 nt transcripts are partially susceptible, and 25mer RNA is fully susceptible to cleavage. Authors proposed that RNA-20 is fully protected in the ternary elongation complex and 21-24mer RNA is conditionally protected depending on the degree of promoter DNA scrunching. Therefore we assume that 9mer RNA as shown in figure 3.9 are only susceptible to decapping when released, and otherwise protected inside RNAP. We thus tested transcript length requirements of NudC to establish a method for capping and decapping without RNAP-template complex disassembly. We designed a template *rnai* 18, which allows for transcripts of 18 nt length (figure 3.10 A) in an experimental setup that would assure to only show elongating RNA. Figure 3.10 B shows a 18 nt NADylated transcript partially resistant to

NudC cleavage, but 23 nt transcript is fully susceptible to decapping. This suggests that NudC and RNAP can come into close contact during decapping.

3.4.7 *NudC purification shows that NudC replaces Zn²⁺ with Fe²⁺*

NudC can be overexpressed from a plasmid and purified via affinity chromatography and N-terminal His-tag. We observed that when purified in buffers without EDTA, highly concentrated NudC (100-300 μM) appears red. This colour can be removed by long dialysis (>24 h) in EDTA, suggesting colouration is conferred by metal binding. Thus we have subjected a red purification sample as well as a colour-reduced sample (24 h dialysis in 10 mM EDTA) to inductively coupled plasma mass spectrometry (ICP-MS), which was performed by Dr Kevin Waldron lab (Newcastle University, ICaMB). The analysis showed that approx. 1 μM of “red” NudC without EDTA contains 0.8 μM Zn²⁺ and 0.3 μM Fe²⁺, while values for Mn²⁺, Co²⁺, Ni²⁺ and Cu²⁺ were 0.01, 0.0015 and 0.031, respectively. Dialysis in EDTA reduced metal to 0.14 μM Fe²⁺ and 0.5 μM Zn²⁺ per μM NudC.

If we assume that NudC concentration was not assessed accurately, we can conclude that one mol of NudC binds one mol Zn²⁺ if available, or replaces it with Fe²⁺, if Zn²⁺ is not available. 0.8 μM Zn²⁺ and 0.3 μM Fe²⁺ would add up to 1.1 μM metal, indicating the assay contained more than the predicted 1 μM NudC (to which metal ions were bound). Alternatively, if the real NudC concentration was 0.8 μM , one mol of NudC coordinates one mol of Zn²⁺; or two mol of NudC coordinate one mol of Fe²⁺. As oxidised Fe²⁺ appears red, the observed colour might have been caused by Fe²⁺ binding in the protein. NudC is active in the cell as homodimer (Höfer *et al.*, 2016). Our “red” NudC purification is active *in vitro*, but we cannot be certain that activity is only conferred by Zn²⁺-bound NudC.

3.4.8 *Mitochondrial RNA polymerase incorporates nucleotide-analogues as initiating nucleotides*

In the years 2016 and 2017 several studies have been published observing NAD⁺-capping in prokaryotes as well as eukaryotes (yeast and mammals). Notably, yeast mitochondrial transcripts were found to be NADylated (Walters *et al.*, 2017). Both bacterial and eukaryotic nuclear RNAPs are multisubunit enzymes. Thus we wanted to test if mitochondrial RNAP, which is a single subunit enzyme, would perform NCIN-capping. We performed *in vitro* transcription experiments with human POLMRT. mtDNA genomes have only two promoters, light and heavy strand promoter (lsp, hsp). Lsp transcription start is A. Figure 3.11 shows POLMRT, assisted by essential transcription factors TFAM and TFB2M, utilises all tested

adenine-containing metabolites. We used NudC to verify presence of the cap and rule out initiation at unspecific start sites. Only FAD-RNA was not cleaved by NudC. Interestingly, mitochondrial transcription shows generally low abortive initiation, except for DP-CoA.

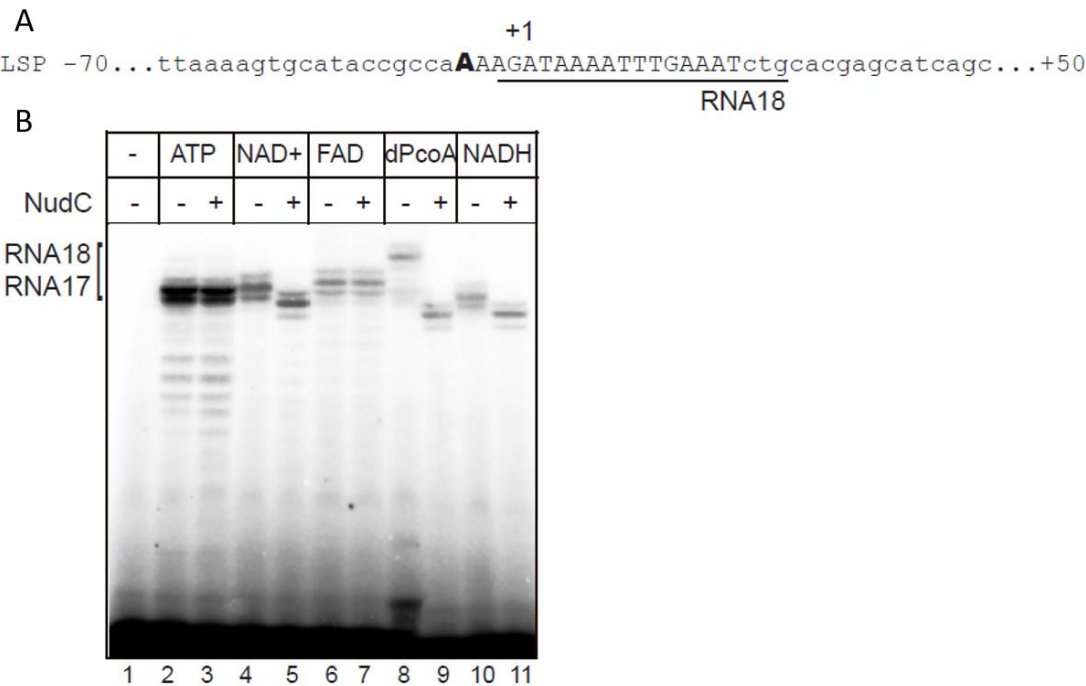


Figure 3.11 NCIN-capping by POLMRT. A. lps is a +1A promoter. If CTP is omitted in the assay, a 18 nt RNA product is produced by POLMRT. B. POLMRT, in association to lps, TFAM and TPB2M, produced 18mer transcripts of varying length in dependence on size of initiating substrate. Metabolite concentration was very high (1 mM) to allow it to compete with ATP (10 μM) substrate. NudC was added to half of the assays to verify presence of cofactor cap (Julius, Riaz-Bradley and Yuzenkova, 2018).

3.5 Discussion

3.5.1 Metabolites and cell wall precursors on 5'-RNA were identified in vivo and shown to be incorporated by RNAP in vitro

Hyptheses about the capping mechanism were controversial at first. Liu group proposed a post-transcriptional mechanism, comparable to classical, eukaryotic capping. However, such mechanism has not been proven to date. Other groups proposed transcription initiation as mechanism, inspired by a publication of Malygin and Shemyakin around 40 years ago, who synthesised dinucleotides from DP-CoA, NAD⁺ or FAD and labelled nucleotide (Malygin and Shemyakin, 1979). Since RNAP catalyses the bond between the 3'-OH of the first NTP and 5'-α-phosphate of the second NTP, the 5'-side of the first substrate is not important for the reaction and another chemical group could theoretically substitute for triphosphate group of

the initiating nucleotide. Here we show that nucleotide-containing small molecules can initiate transcription on a promoter given the nucleotide moiety of initiator base pairs with the template at the transcription start site (TSS or +1) (figure 3.2 and figure 3.4). Adenosine-containing molecules can be used to cap transcripts on +1A promoters, and uridine-containing compounds were able to initiate transcription on +1U promoter. What makes this finding interesting is that these nucleotide-containing substrates are highly abundant functional molecules in the cell, which are known to interact with distinct pathways of metabolism or cell wall synthesis. Thus, their presence in RNA primary structure could potentially add a whole new layer to the function of those RNAs, or cofactors.

In order to draw conclusions about the likelihood of NCIN-capping being performed by RNAP *in vivo*, we measured K_M for production of short transcripts initiated by ATP, NAD^+ , NADH, UTP, UDP-Glc and UDP-GlcNAc *in vitro*. RNAP utilises these substrates at varying efficiency, but at concentrations markedly lower than their cellular concentration during bacterial growth (tables 3.2 and 3.3), in all cases except NADH. NADH is incorporated at similar efficiency to NAD^+ , while its' cellular concentration was determined to be lower (Bennett, Kimball and Gao, 2009).

DP-CoA, FAD, UDP-Glc and UDP-GlcNAc caps have been found on bacterial or eukaryotic RNA *in vivo* (Chen *et al.*, 2009; Kowtoniuk *et al.*, 2009; J. Wang *et al.*, 2019). A number of unidentified 5'-RNA conjugates was found but not further characterised, suggesting that other NCIN-caps exist (Chen *et al.*, 2009; Kowtoniuk *et al.*, 2009; J. Wang *et al.*, 2019). These could be other highly abundant adenine- and uridine-containing molecules, as well as guanosine- or cytidine-containing molecules.

It needs to be stressed that *in vivo*, for any encoded RNA only a small percentage of transcripts are found to be capped. The most frequently capped RNA in *E. coli* was RNA1 with 13 % (Cahová *et al.*, 2015). Hence, the regulatory event behind NCIN-capping might be availability of substrates in competition with canonical nucleotide. Nonetheless, further regulatory elements could be at work such as promoter sequences or transcription factors.

3.5.2 Promoter-mediated regulation of NCIN-capping is controversial

Besides substrate concentration, the template sequence around TSS might take part in capping regulation. The NAD^+ -cap contains besides adenosine another nucleoside-like structure, nicotinamide ribose. Therefore, van der Waals-interactions of this moiety to the -1 base of the template is conceivable. We thus tested the effect of mutation -1 *rnai* template position and

found that overall transcription is affected by identity of -1, but not capping in particular (figure 3.5). Possibly, the degree of interaction of NCIN with -1 could vary between promoters and between NCINs, or regulation could take place at another (not -1) site. From our data we cannot conclude a regulation of NCIN-capping by promoter -1 position.

Other groups further characterised promoter structural elements which regulate capping and hypothesised that -1 base affects NADylation efficiency on *rnaI* promoter and in *B. subtilis* transcription, contrary to our results (Bird *et al.*, 2016; Frindert *et al.*, 2018; Vvedenskaya *et al.*, 2018). A crystal structure of *T. thermophilus* RPo in complex with two initiating NTPs revealed that +1 NTP undergoes stacking interaction with template -1 base which serves to stabilise initiation, if -1 template is a purine base (Basu *et al.*, 2014). Basu *et al.* hypothesise that this stacking interaction explains prevalence of non-template -1 pyrimidine, +1 purine sequences on many natural promoters. Intuitively, initiating metabolites should interfere with such stacking interactions. Vvedenskaya *et al.* show that a promoter sequence of $H_{-3}R_{-2}R_{-1}A_{+1}S_{+2}W_{+3}W_{+4}$, where A is the TSS, increases likelihood of NAD^{+} incorporation 40-fold as opposed to every single one of those positions (except +1) being exchanged by their individual anti-consensus giving the sequence $G_{-3}Y_{-2}Y_{-1}A_{+1}W_{+2}S_{+3}S_{+4}$. The exchange of one individual position to the anti-consensus nucleotide results in 1.2 to 4.1 fold decreased cofactor incorporation, depending on the exact position (Vvedenskaya *et al.*, 2018). Like most promoter elements this proposed consensus is characterised by a low G/C content. *E. coli* promoters are about 5% less GC-rich than the rest of the genome (Meysman *et al.*, 2014). This structural feature facilitates more efficient DNA melting which might be a contributing factor for incorporation of bulky caps. Promoter curvature, as well as downstream template secondary structure can affect promoter binding, initiation efficiency and promoter escape (Kalate, Kulkarni and Nagaraja, 2002; Meysman *et al.*, 2014), which could play in favour or against incorporation of specific caps.

NADylation in *B. subtilis* was shown to underlie a certain degree of regulation by promoter -1 position. In *B. subtilis* transcription, -1 base was mutated for promoter of *veg* gene (Frindert *et al.*, 2018). Veg is a transcription factor that regulates genes for biofilm formation (Lei *et al.*, 2013). Mutation of *veg* promoter -1T to A, C, or G reduced overall transcription efficiency and resulted in lower proportions of NAD^{+} -capped RNA *in vitro*, if both ATP and NAD^{+} were given at equimolar concentration. *In vivo* transcription from *veg* from plasmid DNA in a Δveg strain lead to NADylation of 3.6 % wt *veg* mRNA, or 2.1 % -1C *veg* mRNA (Frindert *et al.*, 2018). Also NAD^{+} -capping in *S. aureus* was found to be sensitive to changes in -1 template position (Morales-Filloy *et al.*, 2019).

All studies on promoter sequence- mediate capping regulation have focussed on NAD⁺ as initiator. Other NCINs, such as the phosphopantetheine of DP-CoA, do not bear a nucleotide-like structure on their 5'-end. Thus van der Waals interaction of these moieties with -1 is unlikely. Future research might aim to identify signals to regulate incorporation of other metabolites. Possibly, different metabolites might interact with promoters in different ways. In *E. coli* and *S. venezuelae* cells, more NADylated RNA (>3000 per cell) as opposed to DP-CoA-capped RNA (50-200 copies per cell) was found (Chen *et al.*, 2009; Kowtoniuk *et al.*, 2009).

Taking the presented results into account we conclude that -1 promoter position may take part in regulation of NCIN-capping on some promoters. However, due to the presented controversy, we do not view -1 position as major regulatory factor.

3.5.3 Protein structural determinants of NCIN-capping were found in σ and β subunits of *EcRNAP*

The structural study of Basu *et al.* showed that σ region 3.2 ($\sigma_{3.2}$) does not interact with initiating NTP (Basu *et al.*, 2014). However, involvement of $\sigma_{3.2}$ in initiation and protrusion of “sigma finger” (a loop of $\sigma_{3.2}$) was shown previously (Kulbachinskiy and Mustaev, 2006; M. S. Paget, 2015). The crystal structure of RNAP in complex with template DNA and NAD⁺pC “dinucleotide” revealed NAD⁺ to be in vicinity to β subunit Rif-pocket residues and $\sigma_{3.2}$ (figure 3.8 B) (Bird *et al.*, 2016; Julius and Yuzenkova, 2017). We suspected a role of *E. coli* $\sigma^{70}_{3.2}$ as well as Rif-pocket residues as possible contributors to capping regulation.

Deletion of $\sigma^{70}_{3.2}$ did not inhibit *in vitro* transcription or capping. Initiation was not affected by $\Delta\sigma^{70}_{3.2}$ for most caps, only the very bulky compound UDP-MurNAc-AA5 was incorporated only in the mutant holoenzyme (figure 3.6 E). Likely, size of the capping molecule is limited sterically by the σ finger loop. This leads us to hypothesise that holoenzymes of alternative σ factors might alter the cap spectrum. Bird *et al.* found that stationary phase cells have higher amounts of NADylated RNA (Bird *et al.*, 2016). We thus tested whether *EcRNAP*- σ^S holoenzyme allows for incorporation of NCINs. Our results do not indicate that σ^S enhances capping of RNA1 (figure 3.6 B and C). This experiment, however, was qualitative and not quantitative. Possibly, the observed increase NADylation in stationary phase of *E. coli* growth could be attributed to changes in RNA degradation, or involvement of other factors. The involvement of transcription factors in NCIN-capping specifically has not been investigated so far.

We further observed that rifampicin-resistant mutants of EcRNAP are deficient in NAD⁺-capping *in vitro*. The amino acid substitutions of these mutants inhibit rifampicin binding either directly or via transmitting conformational changes across the structure (Campbell *et al.*, 2001a). Mutation of residues 513 and 516, in closest vicinity to NAD⁺ in the crystal structure, had most detrimental effect on NAD⁺-capping (figure 3.8 A). Both positions are hydrophilic residues in wt EcRNAP and hydrophobic in mutants, so they might not accommodate NAD⁺ in the catalytic centre, such that transcription is not initiated. Alternatively, the mutations might alter the conformation of the enzyme close to the catalytic site such that accommodation of NMN is less favourable than in the wildtype. Typically, fitness of rifampicin-resistant *E. coli* is reduced (slow growth, increased heat-sensitivity, altered colony morphology) (Brandis *et al.*, 2015). It is interesting to imagine a connection between faults in NCIN-capping and reduced fitness phenotypes of rifampicin-resistant *E. coli*.

In vivo, fitness loss by mutations in the rif pocket can be eased by secondary mutations at another domain of RNAP (Brandis *et al.*, 2012). Resistance can be maintained while fitness loss is compensated with mutations in RNAP subunits α , β , or β' . *In vivo*, also mutations in non-RNAP genes and intergenic regions were identified to be associated with fitness compensation in rifampicin-resistant mutants (Comas *et al.*, 2012). According to Brandis *et al.*, compensatory mutations located to RNAP subunit genes can act via restoring the environment of the wt Rif-pocket, affecting interactions between RNAP and RNA, affecting interactions between RNAP subunits, or affecting interactions between RNAP and NTPs (Brandis *et al.*, 2012). It could be interesting in the future to study the effect of compensatory mutations on RNAP and NCIN-capping. If a strain was found that has NCIN-capping deficiency due to rif-pocket mutation in RNAP without fitness loss, this could allow to study NCIN-cap function *in vivo*.

3.5.4 NCIN-capping stimulates promoter escape

We further observed that capping seems to stabilise the transcript during initiation. NAD⁺-RNA displayed a reduction in abortive product formation as opposed to ATP-initiated transcription *in vitro* (figure 3.7). Addition of low levels of rifampicin to a transcription reaction leads to inhibition of RNA elongation (Campbell *et al.*, 2001a) and promotes release of abortive products. In the case of NAD⁺-initiation abortive synthesis is not enhanced. As NMN was not conclusively shown to base pair with the template, it is unclear whether the initiator “counts as” dinucleotide to RNAP. In that case, the first phosphodiester bond formed

would create an apparent trinucleotide. The ATP-initiated trinucleotide abortive product was less affected by presence of rifampicin as compared to ATP-dinucleotide, arguing that a trinucleotide competes better against rifampicin.

As a functional consequence, enhanced promoter escape might make up for the lower efficiency in NAD⁺-initiation as compared to canonical RNA initiation. In other words, the number of NAD⁺-initiated full-length transcripts in a reality of NAD⁺-enhanced promoter escape, is statistically higher as opposed to an alternative reality in which NAD⁺-initiation was aborted at the same rate as regular initiation. However, there is no evidence to support this theory.

3.5.5 Mitochondrial single-subunit RNAP readily incorporates adenosine-containing metabolites as 5'-caps

NCIN-capping was shown for both bacterial and eukaryotic RNAPs (Cahová *et al.*, 2015; Bird *et al.*, 2016; Jiao *et al.*, 2017; Julius and Yuzenkova, 2017; Walters *et al.*, 2017). Both are multi-subunit RNA Polymerases. We then observed that also single-subunit *H. sapiens* mitochondrial RNAP caps RNA with NAD⁺, NADH, FAD and DP-CoA *in vitro* (figure 3.11). Our findings were supported by *in vivo* data (Walters *et al.*, 2017; Bird *et al.*, 2018). While ATP was a favoured initiator (giving stronger signal on denaturing gel), all tested caps were incorporated (NAD⁺ around 25% as efficiently as ATP, other caps around 10-15%).

We want to stress that the mammalian mitochondrial genome is transcribed from only two promoters, resulting in long polycistronic RNAs which require processing. Thus only the first RNA adjacent to promoter can be capped by POLMRT. As first genes behind promoters *lsp* and *hsp* are rRNA and tRNA, respectively, they will undergo 5'-processing. However, *in vivo*, 10% of *lsp*-transcript were shown to carry NAD⁺ modification (Bird *et al.*, 2018).

In yeast, all NADylated mitochondrial RNAs identified by Walters *et al.* were not 5'-processed (Walters *et al.*, 2017). The yeast mitochondrial genome has 13 promoters, all of which are +1A thus could potentially lead to cofactor capping (Turk *et al.*, 2013). Walters *et al.* found four frequently NADylated transcripts, of F0-ATP synthase subunit c (ATP9), cytochrome c oxidase subunit 2 (COX2) and rRNAs for 21S and 15S ribosomal RNA subunits. Mitochondrial capping was shown very high in yeast, with 50% of mRNA of COX2 RNA being NAD⁺-modified and 10 % NADH-modified (Bird *et al.*, 2018). Other cofactor-caps could exist on mitochondrial RNA as well.

3.5.6 NCIN-caps can be removed by NUDIX hydrolases

The NUDIX hydrolases are pyrophosphatases that act on substrates of general structure NUCleoside DIPhosphate linked to another moiety X (McLennan, 2006). The NUDIX box common to all NUDIX family members comprises the catalytic amino acids sequence $Gx_5Ex_5[UA]xREx_2EExGU$ where the central glutamate residues E174, E177 and E178 (*E. coli* NudC) interact with a divalent cation. NAD^+ hydrolysis to NMN and AMP relies on presence of a Mg^{2+} (Zhang *et al.*, 2016).

A Zn-binding motif serves formation of a NudC homodimer and is essential for protein structure (Zhang *et al.*, 2016). When we purified NudC from overexpression plasmid we observed a red colouration. Metal analysis via ICP-MS showed that our purification contains both Fe^{2+} and Zn^{2+} . We therefore hypothesise that NudC overexpression in complex medium depleted free Zn^{2+} such that many NudC molecules replaced Zn^{2+} with Fe^{2+} ions at the Zn^{2+} binding site. Such phenomenon has been documented in literature. Replacement of Zn^{2+} with other metals can lead to structural or functional deficiencies of the Zn^{2+} binding protein (Deegan *et al.*, 2011; Kluska, Adamczyk and Krężel, 2018). Our “red” NudC purification is active in *in vitro* capping. However, we cannot proclaim that (partially) Fe^{2+} -bound NudC remains active. Possibly, a Zn^{2+} -bound portion of the NudC population confers activity, contaminated by inactive Fe^{2+} -bound NudC.

As can be seen in figure 3.12, NudC homodimer binds one NAD^+ molecule per monomer (Zhang *et al.*, 2016). The C-terminal domain (CTD) binds NAD^+ at the NMN part between three hydrophobic residues I132, W194 and M201. Nicotinamide ribose 3'-OH interacts with S199, while the amide of nicotinamide interacts with Q192 and A241. Adenine base interacts with another hydrophobic pocket at the dimer interface (Höfer *et al.*, 2016). One residue each of the NudC dimer, F160 on NudC molecule A and Y124 on NudC molecule B. Further, 2'-OH of adenine interacts with E111 and Y124 of NudC molecule B. Thus binding of two NAD(H) requires NudC dimerization (Zhang *et al.*, 2016).

Mutation studies of the enzyme show that it interacts with the NAD^+ on RNA and not with the RNA chain (Zhang *et al.*, 2016), while another study describes it as ssRNA-specific, binding RNA non-sequence specific but rather via charge interaction, and favouring NADylated RNA over free NAD^+ substrate (Höfer *et al.*, 2016). We and others have seen that NudC performs also decapping of NADH and DpCoA (figures 3.9 and 3.11) (Bird *et al.*, 2016; Julius and Yuzenkova, 2017; Julius, Riaz-Bradley and Yuzenkova, 2018), while it cleaves FAD

unreliably. Possibly, buffer conditions are responsible for inconclusive results on FAD-decapping.

It was shown that NudC deletion in *E. coli* increases NADylation of RNA1 from 13% to 26% in *E. coli* (Cahová *et al.*, 2015). It would be interesting to investigate whether NudC deletion also affects levels of other metabolite caps. *E. coli* encodes 13 different NUDIX hydrolases (Xu *et al.*, 2006), some of which might participate in metabolite decapping, potentially in a metabolite-specific, or a broadly specific fashion (Julius and Yuzenkova, 2017). One path of research our lab intends to focus on in the future is identification of other cap-specific NUDIX hydrolases in *E. coli*.

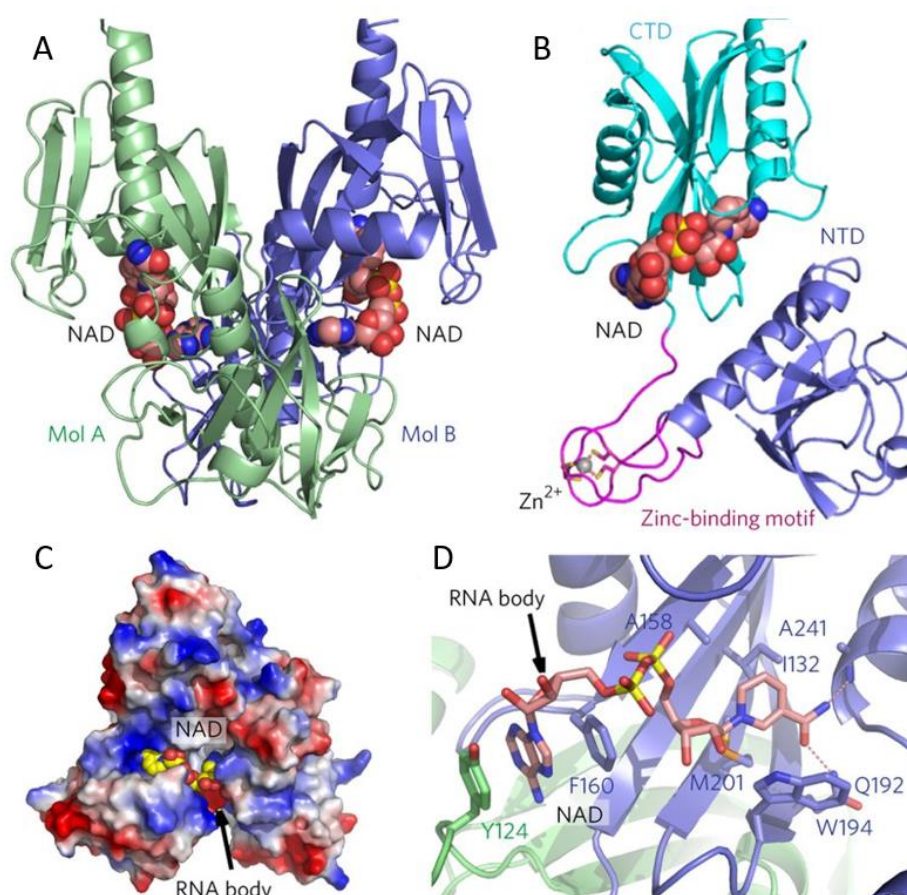


Figure 3.12 Crystal structure of NudC in complex with NAD⁺ from (Höfer *et al.*, 2016). A. NudC homodimer (green and blue ribbons) binds two NAD⁺ (red spheres), one molecule of NAD⁺ per molecule of NudC. B. Different structural regions of one NudC. N-terminal domain (NTD, dark blue), C-terminal domain (CTD, bright blue) binds NAD⁺, and Zn²⁺-binding motif (pink), in complex with one Zn²⁺ (grey sphere), which is important for protein folding. C. Surface of NudC in complex with NAD⁺. Blue areas are positively charged and were proposed to interact with RNA unspecifically at the NAD⁺-RNA extension channel (indicated by arrow). D. NAD⁺-binding is conferred by NudC residues shown as stick models. The location of RNA-extension of NAD⁺ is indicated by the arrow. This view shows that residues from NudC molecule B (blue) and A (green) are involved in substrate binding.

Thus far, NudC has not been found in association to degradosome or other RNA-processing multienzyme complexes. The *E. coli* RNA degradation machinery degradosome is a complex based on scaffolding and endoribonuclease activity of RNase E, in interaction with exoribonuclease PNPase (Polynucleotide phosphohydrolase) which also performs polyadenylation, an RNA helicase and an enolase (Bandyra *et al.*, 2013). RNase E shows strong preference for 5'-monophosphorylated RNA, despite its activity as endonuclease (Mohanty and Kushner, 2016). Thus the degradation pathway of many triphosphorylated RNAs involves a pyrophosphohydrolase step (Silva *et al.*, 2011). In *E. coli* this is performed by RppH, another NUDIX hydrolase. NCIN-capped RNA degradation might rely on processing with NUDIX hydrolases to yield a monophosphorylated 5'-end (Jäschke *et al.*, 2016). Temporary or functional association of RNase E to polyadenylation enzyme poly(A) polymerase (PAP) and RNA chaperone Hfq was documented for RNA processing (Repoila and Darfeuille, 2009). If NudC-decapping was a prerequisite of NAD⁺-RNA degradation, a (temporary) association to an RNA processing or degradation complex would be expected.

In *B. subtilis*, both 5'-triphosphate and 5'- NAD⁺ were found to be cleaved by BsRppH (Frindert *et al.*, 2018). This means that both, NCIN-capped and triphosphorylated RNAs, likely undergo the same route of degradation. *B. subtilis* has no RNase E homolog, its main RNases are RNase Y and RNase J (Bandyra *et al.*, 2013). Further, RNase J1 has important function in mRNA degradation. It has both 5'-exonuclease and endonuclease activity and binds to 5'-mRNA, and can then travel downstream to cleave internal regions (Yao, Sharp and Bechhofer, 2009). In analogy to findings in *E. coli* (Cahová *et al.*, 2015), *B. subtilis* NAD⁺-RNA was protected against RNase J1 cleavage, and required processing by BsRppH (Frindert *et al.*, 2018).

In eukaryotes, NCIN-capping promotes RNA decay (Jiao *et al.*, 2017; Kiledjian, 2018). The classical eukaryotic m⁷G cap is removed by NUDIX hydrolases Dcp2, Nudt3 and Nudt16 *in vivo* and further by Nudt2, Nudt12, Nudt15, Nudt17 and Nudt19 at least *in vitro* (Kiledjian, 2018). Besides NUDIX hydrolases, in eukaryotes also the DXO family is known to function as decapping enzymes with activity both on m⁷G-cap as well as NAD⁺-cap. However, cleavage by DXO removes the complete cap structure as opposed to removal of NMN by NudC (see figure 3.13) (Jiao *et al.*, 2017; Kiledjian, 2018), thus the two enzyme classes yield different products which might have effect on both RNA and metabolite fate. Further, DXO is a 5'-3' exonuclease and was suggested to hydrolyse the transcript following decapping (Jiao *et al.*, 2017). A DXO-equivalent in prokaryotes has not been found to date.

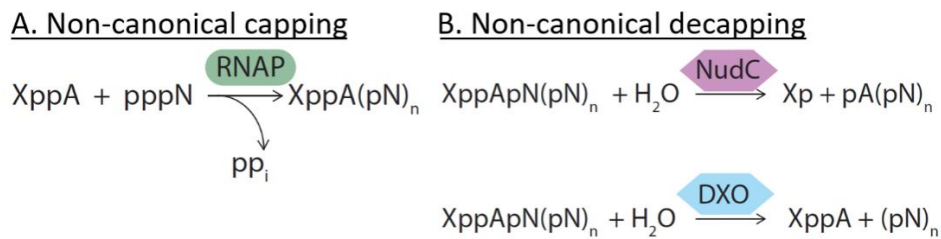


Figure 3.13 Schematic of NCIN-capping and decapping. A. RNAP caps RNA by initiating transcription with an adenosine diphosphate attached to a moiety X (XppA). Phosphodiester bond formation by nucleophilic attack of 3'-OH of XppA on α -phosphate of pppN forms XppApN, pyrophosphate (PPi) is released. B. NCIN-decapping as performed by NudC-like or DXO-like enzymes. NudC hydrolyses the phosphodiester bond between X and nucleotide moiety, yielding monophosphorylated pX and 5'-pA-RNA. DXO hydrolyses the phosphodiester bond between the XppA cap and the penultimate nucleotide, restoring the XppA molecule and monophosphorylated 5'-pN-RNA (Julius and Yuzenkova, 2019).

Another contributing factor to RNA stability was recently discovered in eukaryotes. The first canonical nucleotide of the transcript, so the nucleotide adjacent to the m⁷G-cap, is usually methylated, giving for example a 2'-O-methyladenosine (A_m). But additional methylation giving a N-6,2'-O-dimethyladenosine (m⁶A_m) causes resistance of the cap to decapping NUDIX hydrolase Dcp2, thus conferring increased RNA stability as opposed to single methylated initiating nucleotides (Mauer *et al.*, 2017; Kiledjian, 2018). This methylation is reversible and is considered an epitranscriptomic regulation process. 30% of mammalian cellular RNA carries the “extended mRNA cap” (Mauer *et al.*, 2017). It is possible that also the NCIN-decapping process is amended by modifications on adjoining base or bases. The studies which identified 5'-metabolite caps also detected a variety of nucleotide modifications (Chen *et al.*, 2009; Kowtoniuk *et al.*, 2009).

3.5.7 Metabolite-caps might be incorporated post-transcriptionally

Some findings about NCIN-capping remain controversial. Some NAD⁺-capped RNAs have been found that might not arise from the 5' end of a transcript, such as scaRNAs found in eukaryotes which are generated by processing of RNA (Kiledjian, 2018). We should consider the possibility that alternative pathways exist to cap RNA post-transcriptionally (Chen *et al.*, 2009). In cellular NAD(H) *de novo* biosynthesis, a reverse reaction to NudC exists, performed by nicotinate mononucleotide (NaMN) adenylyltransferase NadD. In NAD⁺ synthesis, NadD catalyses the conversion of NaMN to nicotinate-adenine dinucleotide, which is then aminated to form NAD⁺ (Han and Eiteman, 2018). But NadD also performs an alternative reaction, the conversion of NMN to NAD⁺ (reverse to NudC) (Han and Eiteman, 2018) (see figure 3.14). Via such a pathway, NadD might theoretically be active in the process of RNA NADylation,

which we are going to investigate further. The NAD biosynthesis pathway is similar between eukaryotes and prokaryotes, but eukaryotic NadD does not differentiate between NaMN and NMN substrates, while bacterial NadD favours NaMN substrate (Wang *et al.*, 2017).

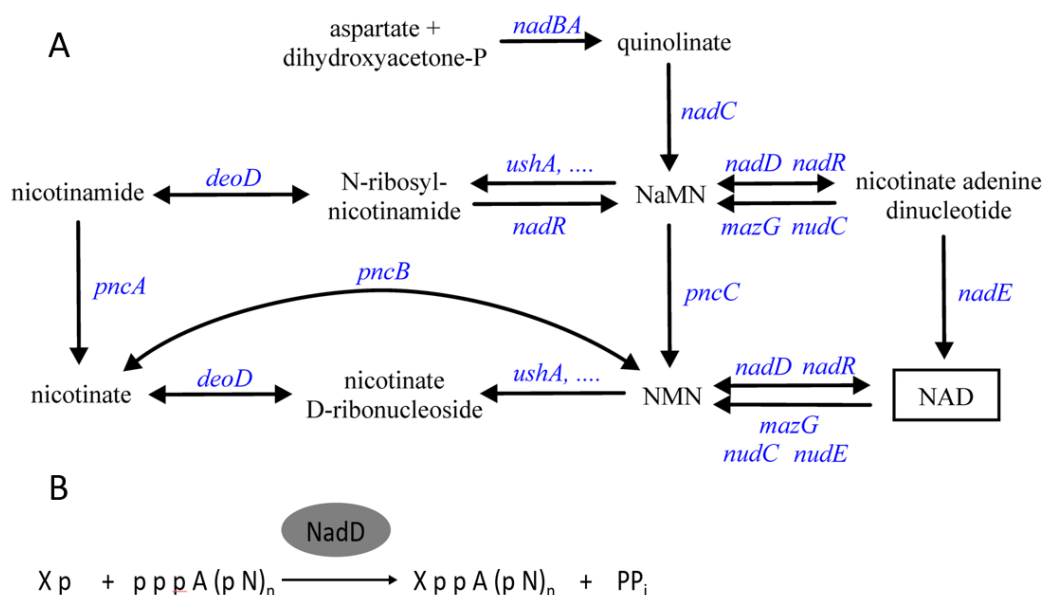


Figure 3.14 Theory of a post-transcriptionally active metabolite-capping enzyme. A. *De novo* synthesis of NAD⁺ in *E. coli*, from (Han and Eiteman, 2018). Enzyme gene names are given in blue, NAD⁺-precursors in black, arrows show direction of the performed reaction. NaMN adenylyltransferase converts NaMN (nicotinate mononucleotide) to nicotinate adenine dinucleotide which is then aminated by NadE to form NAD⁺. Further, NadD converts NMN to NAD⁺, the reverse reaction to NudC. B. Schematic of NCIN-capping by an adenylyltransferase such as NadD.

In 2000, a study was published that attempted to produce DP-CoA, FAD and NAD⁺ via ribozymatic activity, as it might have happened in a time before protein enzymes (Huang, Bugg and Yarus, 2000). A phosphorylated cofactor precursor Xp (e.g. nicotinamide monophosphate, Flavin monophosphate or 4'-phosphopanthetheine) reacts with a triphosphorylated transcript that starts with ATP (pppApN(pN)_n). Pyrophosphohydrolysis might provide the energy to create the XppApn(pN)_n product, catalysed by a ribozyme. Such a ribozyme has not been found in nature, but might once have existed as part of the RNA world theory (White, 1976; Jeffares, Poole and Penny, 1998).

3.5.8 NCIN-capping might have various physiological implications

Alterations of the RNA primary structure must have downstream effects. If the NCIN-cap is not useful to the cell it might be removed to reconstitute functional RNA, or it might act as signal for destruction of a faulty copy. If the cap has a cellular function, this will be revealed in future research. The RNA 5'-end has special importance in both prokaryotes and

eukaryotes. The role of the classical 5'-cap in nuclear transport, translation initiation, and RNA stability is quite essential for mRNAs. However, classical cap-independent pathways of translation are known for eukaryotes (Mitchell and Parker, 2015). Also in bacteria, the 5'-end could affect translation, in the case of leaderless transcripts, or translation regulation via sRNAs. Further regulatory pathways rely on the 5'-end of RNA, such as riboswitches. Table 3.4 sums up cellular functions of frequently NADylated RNAs that have been found to date. The following part of the discussion deals with the different possible physiological outcomes of NCIN-capping. The mentioned aspects are also discussed in (Julius and Yuzenkova, 2019).

While writing this thesis a first publication describing a functional impact of NADylation of a bacterial RNA has shed a fleck of light on what we might expect to find more of in the future. *S. aureus* encoded RNAIII, sRNA involved in regulation of quorum sensing and virulence, was found to be highly NADylated (36.2% in wildtype cells) (Morales-Filloy *et al.*, 2019). The NAD⁺-bearing population of RNAIII was varied *in vivo*, to determine the physiological impact of level of NAD⁺-capping. For this, Morales-Filloy *et al.* deleted the promoter of RNAIII and recovered mutants with two different plasmids that lead to production of 9.8 % or 24.9 % of NADylated RNAIII. The strain that expressed more NAD⁺-RNAIII was less virulent via decreased expression of alpha-endotoxin Hla. Surprisingly, NADylation was not found to affect secondary/tertiary structure formation of RNAIII, thus the effect of NAD⁺-RNAIII was not conferred by alternative folding of this sRNA. Morales-Filloy *et al.* hypothesise further that the effect of RNAIII NADylation on *S. aureus* virulence represents a link between redox state and virulence of the bacterium.

RNAIII is part of the *agr* quorum sensing system. High cell density is registered by the autoinducer peptide API, which binds the histidine kinase ArgC. ArgC activates AgrA by forming AgrA~P. In one pathway, AgrA~P activates modulins Psm α and Psm β virulence factors which act by disrupting the mammalian cell membrane. On another path, AgrA~P acts as transcription activator on promoters P2 and P3. P2 regulates an operon that encodes Agr proteins AgrA, B, C, and D. The polycistronic transcripts of the *agr* genes is called RNAII. Transcription from P3 gives RNAIII, as well as another toxin, Hld. RNAIII itself has several functions. It negatively regulates expression of Rot, transcriptional repressor of a group of *S. aureus* toxins, and of several surface proteins e.g. coagulase. However, RNAIII also activates production of haemolytic exotoxins Hla and Hld, the first of which was focus of the study (Morales-Filloy *et al.*, 2019). The functioning of sRNAs is further described in the next paragraph.

Table 3.4 Functional destinations of RNAs which have been identified as frequently NAD⁺-capped

Organism	Identified gene functions	Publication
<i>E. coli</i>	sRNAs: plasmid replication, stress response mRNAs: enzymes of metabolism, enzymes of stress response	Cahová <i>et al.</i> 2015
<i>S. cerevisiae</i>	mRNAs: cytoplasmic translation, protein transport/localisation to mitochondria, mitochondrial transport, cellular macromolecule assembly, mitochondrial functions (F0 ATP synthase subunit c, cytochrome c oxidase subunit 2), mitochondrial rRNAs	Walter <i>et al.</i> 2017
HEK293T (<i>H. sapiens</i>)	snRNAs, snoRNAs, scaRNAs, mRNAs (study did not focus on their cellular function)	Jiao <i>et al.</i> 2017
<i>B. subtilis</i>	mRNAs: genes involved in phosphatase activity, nucleotidyltransferase activity, proteolysis, zinc binding, symporter activity, DNA binding and replication, oxidation/reduction, sporulation, transcription regulation, membrane integration, nucleotide binding	Frindert <i>et al.</i> 2018
<i>Arabidopsis</i>	mRNAs of photosynthesis, protein synthesis, cytokinin response, stress response	Wang <i>et al.</i> 2019
<i>S. aureus</i>	sRNA RNAIII: regulation of quorum sensing, virulence	Morales-Fillooy <i>et al.</i> 2019

sRNAs and RNA stability

Bacterial NAD⁺-cap was associated to increased RNA stability via protection from RNase E and RppH (Bird *et al.*, 2016; Jaschke *et al.*, 2016), while eukaryotic NADylation of transcripts was described to promote degradation (Jiao *et al.*, 2017). We think that although NCIN-capping in bacteria might protect the RNA from RppH and RNase E cleavage, this does not mean that stability is increased. In fact, among sRNAs which are highly NADylated in *E. coli*, no apparent correlation between capping and half-life can be observed (see table 3.5). NAD⁺-RNA breakdown might lead over a different route of RNA breakdown. We think that findings that among highly NADylated mRNAs only 5'-fragments of <200 nt were found (Cahová *et al.*, 2015), might be an argument in favour of degradation or endonucleolytic processing of NADylated mRNAs.

RNA half-life can be very variable with 3-100 min in yeast and usually less than 10 min in bacteria (Silva *et al.*, 2011), dependent on type of RNA and growth state (Deutscher, 2003).

Stationary phase RNA was shown to be more frequently NADylated (Bird *et al.*, 2016). If this is evidence for a direct effect of 5'-NAD⁺ stability, or if other factors upregulate NCIN-capping during stationary growth, is unknown. RNA half-life can have a strong downstream effect. A more stable mRNA will be translated more often before its degradation, and a more stable sRNA could regulate more genes. Table 3.5 sums up known half-lives, interaction partners and functions of sRNAs that were shown to be highly NADylated in *E. coli*. Half-life of the most frequently capped sRNAs seems to be quite diverse. Thus on first sight a relationship between sRNA capping and stability is not apparent.

sRNAs are regulatory RNAs that bind to a target sequence. Few bind to proteins, e.g. 6S RNA sequesters growth phase RNAP- σ^{70} by binding σ^{70} . This way, 6S RNA thus regulates stationary phase gene expression by suppressing σ^{70} , but not σ^S (Waters and Storz, 2009). Most sRNAs interact with mRNAs. They may act as translational silencers, binding and occluding the ribosome-binding site (RBS) or other regions in the 5'-untranslated region (UTR) (Waters and Storz, 2009).

Cis-acting sRNAs are encoded on the anti-sense strand of their target mRNA and have complete sequence homology, thus they are specific to their particular mRNA. Once produced, the sRNA can bind to the target but can also diffuse freely. This type of regulation is often found to control copy number of plasmids or transposons or in toxin/anti-toxin systems (Waters and Storz, 2009). One example for the *cis* sRNA is RNA1, the most highly NADylated RNA in *E. coli* (Cahová *et al.*, 2015). It regulates the replication of plasmid ColE1 by sequestering the ColE1 primer (called RNA2) (Tamm and Polisky, 1985).

Trans-acting sRNAs are generally transcribed from another location than their target mRNA. The sequence homology is limited and many *trans* sRNAs have several target RNAs. Therefore, RNA chaperone Hfq frequently mediates between sRNA and target mRNA. One example of a *trans* sRNA is GcvB, one of the sRNAs previously identified to be highly NADylated *in vivo* (table 3.5) (Cahová *et al.*, 2015). It binds mRNA upstream of the translation start site. Formation of dsRNA then leads to degradation (Waters and Storz, 2009).

Table 3.5 sRNAs which were found to be highly NADylated in *E.coli*, including information on half-life, RNA and protein binding partners, and function (Julius and Yuzenkova, 2019).

sRNA	cis trans	or	target mRNA(s)	protein binding	half-life [min]	function	References
RNA1	trans		RNA2	Rop	1	ColE1 replication control	(Tamm and Polisky, 1985; Bird <i>et al.</i> , 2016)
GadY	trans/cis*		GadX, GDS	Hfq/Hfq-independent*	n.d.	pH stress response	(Opdyke <i>et al.</i> , 2011; Cahová <i>et al.</i> , 2015; Negrete and Shiloach, 2015; Bird <i>et al.</i> , 2016)
CopA	trans		CopT	n.d.	3	r1 plasmid replication	(Gerhard E, Wagner H, 1986; Cahová <i>et al.</i> , 2015)
GcvB	trans		CycA	CsrA	2	peptide/aa transport, glycine transport	(Pulvermacher, Stauffer and Stauffer, 2009; Stauffer and Stauffer, 2012; Cahová <i>et al.</i> , 2015)
ChiX	trans		ChiP	Hfq	27	regulation of outer membrane channel	(Mandin and Gottesman, 2009; Edwards <i>et al.</i> , 2011; Cahová <i>et al.</i> , 2015)
McaS	trans		YdeH, YdcT	Hfq, CsrA	20	biofilm regulation	(Jørgensen <i>et al.</i> , 2013; Cahová <i>et al.</i> , 2015)
GlmY	trans		GlmZ	PAPI	1.4	peptidoglycan synthesis pathway	(Reichenbach <i>et al.</i> , 2008; Cahová <i>et al.</i> , 2015)
DsrA	trans		RpoS, Hns, MreB, RbsD	Hfq	23	transcription, cell wall sythesis, ribose metabolism	(Cayrol <i>et al.</i> , 2009; Cahová <i>et al.</i> , 2015; Wu <i>et al.</i> , 2017)

*conflicting documentations, n.d. = not determined

More rarely, sRNA binding activates gene expression, by resolving secondary mRNA structure that inhibits ribosome binding (Repoila and Darfeuille, 2009; Waters and Storz, 2009). The *S. aureus* sRNA RNAIII mentioned earlier is an example of a trans sRNA which acts both as activator and inhibitor of translation. NADylation of RNAIII was found to modulate production of alpha-toxin (Morales-Fillooy *et al.*, 2019). *Trans* sRNAs are often regulated by a specific transcription factor in response to the environmental conditions (e.g. GcvB transcription factor GcvA is part of elevated glycine response), and they pass on this response by translation control of their various targets (Waters and Storz, 2009). If NCIN-capping was to affect the activity of sRNA this could have an accumulative downstream effect.

Translation initiation

The eukaryotic m⁷G-cap is an important structural feature of assembly of the translation initiation complex (Ramanathan, Robb and Chan, 2016). Experiments with luciferase reporter gene in cell culture have revealed that, in eukaryotes, presence of NAD⁺-cap inhibits translation initiation (Jiao *et al.*, 2017). However, as m⁷G-capping is a post-transcriptional process, the m⁷G-cap could be attached to mRNA after NCIN-decapping by NUDIX hydrolases or DXO.

The prokaryotic translation machinery does not rely on a 5'-cap. On mRNAs with ribosomal binding sites downstream of TSS, translation initiation relies on RNA sequence and secondary structure, and presence of initiation factors IF2 and IF3. On leaderless RNA, the AUG codon is necessary and sufficient recognition signal for ribosome binding. Leaderless translation is initiated by a complex of ribosome, IF2 and ^fMet-tRNA (Moll *et al.*, 2002). NCIN-capping of leaderless RNA would disrupt the start codon (i.e. NAD⁺-UG instead of AUG) and could potentially fail ribosome binding and translation initiation. Alternatively, NCIN-capped translation initiation may rely on the presence of additional factors.

The majority of NADylated RNAs in *E. coli* are sRNAs, thus non-translated. However, many bacteria contain high percentages of leaderless transcripts (e.g. one third of mycobacterial transcripts are leaderless (Shell *et al.*, 2015)). Hence, NCIN-capping of leaderless RNA likely occurs *in vivo*.

Riboswitches and other regulatory 5'-elements

Many regulatory elements are located in the 5'-UTR of bacterial RNA. This region can assimilate a secondary structure in response to environmental stimuli or can bind environmental signals and thus modulate translation (Waters and Storz, 2009). As an example, “RNA thermometers” exhibit temperature-sensitive secondary structure.

Riboswitches are *cis*-acting regulatory elements in the 5'-UTR that bind ligands. Often, they regulate expression of genes for regulation or utilisation of that ligand, as a type of feedback regulation (Waters and Storz, 2009). In riboswitch secondary structures the 5'-terminus can be included, or may remain single stranded (Montange and Batey, 2008). It would be interesting to see the effect of NADylation on riboswitch folding or ligand binding. However, folding of RNAIII in *S. aureus* was not found to be affected by NADylation (Morales-Fillooy *et al.*, 2019).

RNA localisation and RNA-protein interactions

Eukaryotic RNA localisation is often spatially confined by organelles such as nucleus or endoplasmic reticulum. In bacteria, RNA subcellular localisation is mainly undefined. For a long time it was thought that mRNA co-localises with the bacterial nucleoid, because transcription and translation would be coupled. However, more recent findings of fluorescence-microscopy showed that indeed only 4% of translation processes are coupled and co-localised with transcription (Buskilay, Kannaiah and Amster-Choder, 2014). As it seems, mRNA also localises to non-enclosed compartments or subcellular locations which are relevant for their cognate protein. Five patterns of mRNA localisation have been identified. i. Transcription site of chromosome, limited cytoplasmic diffusion ii. Cytoplasmic distribution, iii. Membrane localisation/ cell periphery, iv. Polar vs. septal (Buskilay, Kannaiah and Amster-Choder, 2014).

The mechanism of mRNA localisation is however unknown. It has been proposed that due to bulkiness and RNA-protein interactions simple diffusion through the cytoplasm is unlikely the only mode of transport for RNAs (Buskilay, Kannaiah and Amster-Choder, 2014).

Many protein enzymes involved in metabolism interact with the adenine-containing cofactors like NAD⁺, NADH, FAD, DP-CoA. In addition to that, metabolism enzymes are known that “moonlight” as RNA-binding proteins (Castello, Hentze and Preiss, 2015). RNA interactome studies in yeast revealed 23 enzymes cross-linked to polyadenylated RNA. More than half of them also bind mono-or dinucleotides. Here, an interaction with NAD(H) or NADP(H) is commonly reported. One example is glyceraldehyde-3-phosphate dehydrogenase (GAPDH) of glycolysis. It converts glyceraldehyde-3-phosphate to D-glycerate-1,3-bisphosphate while generating NADH. Besides this, roles in transcription, DNA repair and interferon response have been documented for GAPDH. NAD⁺ and RNA seem to compete for the same dinucleotide-binding site. The presence of NAD⁺ or NADH interferes with GAPDH RNA

binding activity, as a means of regulating which function the enzyme assumes (Castello, Hentze and Preiss, 2015).

Some bacterial proteins have been shown to be attracted by a specific membrane curvature (which differs at body and poles for non-spherical bacteria). Others recognise specific lipids, like cardiolipid, which populates pole membranes more densely in *E. coli*. In eukaryotes, RNA is believed to display *cis*-acting localising elements, or “zip-codes” in untranslated regions. Their mechanism of action is as of yet unresolved (Buskilay, Kannaiah and Amster-Choder, 2014). For prokaryotes it has been hypothesized that “zip-codes” are positioned in coding regions. (Buskilay, Kannaiah and Amster-Choder, 2014). The case of RNA capped with cell wall precursors (Julius and Yuzenkova, 2017) is especially interesting. UDP-Glc and UDP-GlcNAc are present at high concentrations in the cell and are turned over frequently during sugar metabolism and cell wall recycling (Konopka, 2012). It is possible that their incorporation into RNA “tags” the transcript towards the inner membrane. GlcNAc was besides being a precursor for murein, also associated to other cellular functions, such as signalling (Konopka, 2012). In *Streptomyces* it is involved in signalling for sporulation (Konopka, 2012). UDP-Glc- and UDP-GlcNAc-RNAs have been found *in vivo* (J. Wang *et al.*, 2019), though their cellular function is as of yet undefined.

Effect of capping on cofactor pool and cellular redox potential

NCIN-capping might reflect the metabolic state of the cell (Bird *et al.*, 2018; Frindert *et al.*, 2018; Julius, Riaz-Bradley and Yuzenkova, 2018). NAD(H)-capping in mitochondria was shown to be responsive to NAD⁺/NADH levels and ratio (Bird *et al.*, 2018; Grudzien-Nogalska *et al.*, 2018).

The bacterial cytosol is packed with metabolites. Altogether about 300 mM of diverse small molecules that are involved in metabolism can be detected in growing *E. coli* (Bennett, Kimball and Gao, 2009). Proportionally the largest group of these are the amino acids, followed by nucleotides ATP (9.6 mM), and UTP (8.3 mM) (Bennett, Kimball and Gao, 2009). Many metabolic molecules contain nucleoside moieties which would make them potential NCIN-caps. Adenine-containing cofactors of metabolism are present in high concentrations, such as NAD⁺ (2.6 mM), FAD (1.7 mM), Coenzyme A (1.4 mM) or NADP (1.2 mM). Much more highly concentrated are uridine-containing intermediates of sugar metabolism and cell wall synthesis, such as UDP-GlcNAc (9.2 mM), UDP-Glc (2.5 mM) or UDP-glucaronate (5.7 mM) (Bennett, Kimball and Gao, 2009). The intracellular concentrations of these molecules exceeds the previously measured K_M of EcRNAP (Julius

and Yuzenkova, 2017) by far: K_M of ATP utilisation is 0.09 mM, of adenine-cofactors NAD^+ and NADH 0.36 mM and 0.38 mM, respectively. RNAP K_M to UTP is 0.12 mM in comparison to UDP-Glc (0.33 mM) and UDP-GlcNAc (0.3 mM).

NCIN-capping was suggested to be a means of epigenetic regulation in eukaryotes (Kiledjian, 2018), and it could also contribute to an analogous regulation in bacteria. Epigenetics relies on DNA modification and tertiary/quaternary structure affecting the transcriptome. In bacteria, DNA spatial organisation, though not as packed, is likely to affect RNA capping due to molecular crowding and concentration bias. In *S. aureus* it was recently suggested that pathogenicity might be modulated via the bacterium's metabolic or redox state via NAD^+ -capping (Morales-Filloy *et al.*, 2019).

Inversely, NCIN-capping might modulate the pool of free metabolites. This might be most impactful in mitochondria. Mitochondrial ATP-synthesis relies on availability of NAD(H) as redox factor. Therefore, regulation of RNA NADylation might be a crucial event in the response of mitochondria to nutrient or redox stress.

3.6 Conclusion and Outlook

NCIN-caps were verified to be a universal RNA modification or extension of primary structure. We (and others) have shown that NCIN-caps are incorporated into RNA by transcription initiation. Several protein and promoter structural features have been described to likely regulate NCIN-capping. Yet, the actual functional importance of this event remains to be determined. The finding that *S. aureus* alpha-toxin is regulated by the level of NADylated RNAIII is a first sign of functional significance (Morales-Filloy *et al.*, 2019). Besides regulation of virulence factors, NCIN-capping might also direct interaction of the transcript with other molecules, balance free metabolite availability, or regulate gene expression via transcript stability, translation regulation, or RNA localisation. NUDIX and DXO enzymes direct NCIN-RNA degradation.

Future research will show which NCIN-caps exist in nature, how the cap spectrum is regulated, whether NCIN-caps are specifically recognised by proteins or nucleic acids *in vivo*, if there are RNAP-independent pathways of non-canonical capping, and which physiological consequences arise from NCIN-capping and capping deficiency

4 Chapter II: Primase utilises cellular cofactors to initiate primer synthesis

4.2 Introduction

Primase is a specialised ssDNA-dependent RNA polymerase that initiates replication of the genome by producing RNA primer. Since DNA polymerases are inherently unable to initiate *de novo* DNA synthesis, they rely on primer as start points for replication. All bacteria and eukaryotes utilise RNA primer. Among viruses of eukaryotes other approaches to priming are known, such as use of tRNA or even protein primers (Kuchta and Stengel, 2010). Bacterial primase (as well as phage and plasmid primase) is a one subunit enzyme which associates temporarily to a multi-enzyme replisome, and interacts predominantly with helicase. Some bacteriophage primases instead have a helicase subunit (Kuchta and Stengel, 2010).

Eukaryotic and archael primases have two subunits, the catalytic subunit P49 (also called P48 or PriS) and a larger subunit P58 (or PriL) which is involved in template and NTP binding. P49-P58 complex associates to DNA Polymerase α (Pol α) to form the Pol α -primase or pol-prim complex for RNA-DNA primer synthesis (Muzi-Falconi *et al.*, 2003). In mitochondria, interestingly, RNA primer is generated by the transcriptase POLMRT as side product of transcription (Wanrooij *et al.*, 2008). A second primase which generates deoxyribonucleotide primer is known in both eukaryotic nuclei and mitochondria, and is called PRIMPOL (García-Gómez *et al.*, 2013).

The Primer is a short lived product (Balakrishnan and Bambara, 2013). Yet its' synthesis and removal are the rate limiting steps of replication (Bailey, Wing and Steitz, 2006). After providing a 3'-OH to processive DNA polymerase for synthesis of leading strand or Okazaki fragments (DNA stretches between two primers on the lagging strand), primer is removed to facilitate production of a complete DNA strand. Bacteria remove primer by RNase H, a ribonucleic acid endonuclease with specificity for RNA-DNA hybrids, and/or by DNA polymerase I (Pol I) 5'-nuclease activity (Berkower, Leis and Hurwitz, 1973). Pol I fills up the gap between two Okazaki fragments with DNA, and DNA ligase seals the backbone nicks. Eukaryotic primer is removed by one or several flap-endonucleases after being displaced by DNA polymerase (Rossi and Bambara, 2006).

We show that *E. coli* primase DnaG efficiently initiates primer synthesis using nucleotide-containing metabolites such as NADH and FAD as non-canonical initiating substrates (NCIS). This is regulated by residues of DnaG basic ridge amino acids and does not respond

to alterations of primase recognition site -1 position on DNA. We found that stringent response alarmone ppGpp likely inhibits metabolite initiation. We hypothesise that non-canonically initiated primer may affect replication via altered Okazaki fragment processing in bacteria, as Pol I is differentially affected by NCIS. Further, we show that human primase catalytic subunit P49 can initiate primer with methylated guanosine triphosphate (m⁷GTP).

4.2.1 *DnaG is an RNA polymerase that looks like a topoisomerase*

The chapter will use *E. coli* protein nomenclature, unless otherwise stated. Bacterial primases are widely conserved. The DnaG protein forms three distinct domains. The N-terminus bears a zinc-binding domain (ZBD). The Zn²⁺-finger found here confers binding to DNA (Akabayov *et al.*, 2009; Lee *et al.*, 2012; Ilic *et al.*, 2018). The C-terminus is referred to as helicase binding domain (HBD) or helicase interaction domain (HID) (see figure 4.1 A) (Tougu, Peng and Marians, 1994; Kuchta and Stengel, 2010; Ilic *et al.*, 2018). The largest domain in between both peptide termini is the RNA polymerase domain (RPD) containing a TOPRIM fold. The TOPRIM fold is common to many bacterial enzymes that are involved in DNA maintenance such as topoisomerases and recombinases (Rymer *et al.*, 2012; Hou, Biswas and Tsodikov, 2018). Three divalent cations (Mg²⁺, Mn²⁺, or Fe²⁺) are coordinated by conserved residues in the TOPRIM domain for RNA synthesis and NTP binding. Potentially, DnaG oligomerisation (dimer up to hexamer) is important for enzyme activity, as found for T4 primase (Yang *et al.*, 2005; Ilic *et al.*, 2018) and allows for sharing of catalytic sites (Yang *et al.*, 2005). The TOPRIM fold forms the walls of a central cleft in DnaG where most inhibitors as well as cofactors bind (Stamford *et al.*, 1992; Aravind, Leipe and Koonin, 1998; Godson *et al.*, 2000; Keck *et al.*, 2000; Rymer *et al.*, 2012). The N-terminal region of RPD domain also interacts with DNA (Hou, Biswas and Tsodikov, 2018).

Possibly, DnaG travels dsDNA via interaction with RPD N-terminus until a replisome (multi-protein replication complex) is encountered. It binds only one of the two strands but has similar affinity to dsDNA as to ssDNA (Hou, Biswas and Tsodikov, 2018). DnaG recognises a triplet of the consensus 5'-PyrPyrPur (two pyrimidines, one purine) on the non-template strand, thus primer synthesis can potentially be initiated frequently on many genomic locations. Its' preferred recognition sequence is 5'-CTG (Yoda *et al.*, 1988; Yoda and Okazaki, 1991). +1 position (first base pair) is the central pyrimidine of this sequence, thus the initiating nucleotide of the primer has to be a purine (A or G). Variation of -1 base affects initiation efficiency (G>T>C>A).

Figure 4.1 shows important structural information on DnaG and interaction with substrates and template. DnaG catalysis depends on two lysine residues and an EGxxD motif in the catalytic site for phosphodiester bond formation of the first dinucleotide (Lee and Richardson, 2001; Rodina and Godson, 2006). A cluster of acidic residues that bind three divalent cations (Godson *et al.*, 2000; Lee and Richardson, 2005; Hou, Biswas and Tsodikov, 2018), and positively charged residues of the “basic ridge” were found to make contact to the triphosphate moiety of substrates (Keck *et al.*, 2000; Rodina and Godson, 2006). ZBD binds to ssDNA at the recognition site (Yoda and Okazaki, 1991; Swart and Griep, 1993; Hou, Biswas and Tsodikov, 2018).

Two nucleotides are bound within DnaG to form the first phosphodiester bond. This is considered the rate-limiting step of primer synthesis, because the first dinucleotide is not stabilised in the catalytic centre by base pairing of a nascent RNA 3'-end with the template. Thus primer initiation widely depends on substrate concentration. DnaG catalysis functions using three metal ions. This is a main distinction between bacterial primase and other polymerases (including all known transcriptases), which all function by a two metal ion mechanism. The growing RNA chain is displaced as primase travels along the DNA template, while 3'-end of primer remains in the catalytic site between metal binding centre and basic ridge (Hou, Biswas and Tsodikov, 2018). Nucleophilic attack of 3'-OH of RNA to incoming NTP is catalysed by metal ion A (Hou, Biswas and Tsodikov, 2018). Metal ions B and C together with two arginine and one leucine residue were shown to coordinate the triphosphate of the NTP substrate (Rymer *et al.*, 2012) (figure 4.1.1 D). Likely, metals A and B reside within the active site while metal C is introduced by incoming NTP which it serves to position correctly in the active site (Rymer *et al.*, 2012). Metal B aids dissociation of pyrophosphate as the phosphodiester bond is made (Rymer *et al.*, 2012).

After primer synthesis, β clamp is loaded onto the RNA:DNA hybrid, and Pol III is recruited to extend the primer by DNA polymerisation onto the final 3'-OH (Lovett, 2007; Lewis, Jergic and Dixon, 2016). The primase-to-polymerase switch at the primer is coordinated by the clamp loading complex. Clamp loader protein χ binds SSB, breaking the connection of DnaG to SSB (Naue *et al.*, 2013), reducing DnaG stability on DNA. Clamp loader can then displace DnaG and bind itself to load β clamp, which binds Pol III (Yuzhakov, Kelman and O'Donnell, 1999). β clamps have an additional role in recruiting DNA modifying and repair enzymes to the daughter strand (Lewis, Jergic and Dixon, 2016).

After Okazaki fragment synthesis, the primer needs to be removed to facilitate completion of the DNA strand. This is done by RNase H endonucleolytic cleavage of RNA part of RNA-DNA hybrid, and/or Pol I 5'-3'-nuclease activity (Xu, Grindley and Joyce, 2000; Sulej *et al.*, 2012), as described later.

4.2.2 RNA primer synthesis by DnaG continually initiates DNA synthesis

Bacterial replication is initiated at the origin of replication (chromosomal origin, *oriC*) by master initiator protein DnaA, an ATPase-containing enzyme that unwinds and melts *oriC* DNA, creating a replication fork (Jacob, Brenner and Cuzin, 1963; Skarstad and Boye, 1994). After generation of the first primers on *oriC* by DnaG, replication is performed by the bacterial replisome. The replisome is a multi-enzyme complex for DNA replication that moves along DNA to replicate both DNA strands simultaneously in one replication fork, the leading strand in a continuous, and the lagging strand in a discontinuous fashion (Yoshikawa, 1970). DnaG synthesises about 2000-4000 primer per *E. coli* chromosome on the lagging strand (Kusakabe and Richardson, 1997).

Members of the bacterial replisome are main DNA Polymerase III (Pol III), β_2 sliding clamps, clamp loading complex, DNA helicase (DnaB), single strand binding proteins (SSB), and, transiently, primase DnaG and primer displacement polymerase Pol I (Wu *et al.* 1992) (see figure 4.2). The Pol III multisubunit enzyme consists of catalytic subunit α (*dnaE*), 3'-5'-proofreading subunit ϵ (*dnaA*), and subunit θ (*holE*). Three Pol III cores are physically held in the replisome by interaction with τ proteins in clamp loader complex, and interact with β_2 (*dnaN*) sliding clamps on their respective template strands. The clamp loading complex can have different conformations of τ or γ (*dnaX*) and δ (*holA*) and δ' (*holB*): $\tau_3\delta\delta'$, $\gamma_3\delta\delta'$, $\gamma_2\tau\delta\delta'$, or $\tau_2\gamma\delta\delta'$ with τ being the full gene *dnaX* product, and γ being a C-terminally truncated version of DnaX created by translational frameshift or proteolysis (Lovett 2007, Johnson & O'Donnell 2003). The C-terminus of τ interacts with Pol III α as well as DnaB. DnaB is loaded onto ssDNA by helicase loader DnaC (not a stable member of the replisome) (Arias-Palomo *et al.*, 2013). DnaB hexamer sits on lagging strand and unwinds dsDNA at the edge of the replication fork. Replication is initiated as primase DnaG travels along DNA and finds an open replication fork (Hou, Biswas and Tsodikov, 2018). It is also attracted by a recognition sequence on ssDNA (Swart and Griep, 1993). Roughly three DnaG proteins associate with a DnaB hexamer (Arias-Palomo *et al.*, 2013). DnaG is able to synthesise primer without DnaB stimulation, but binding DnaB strongly enhances activity (Johnson, Bhattacharyya and Griep, 2000).

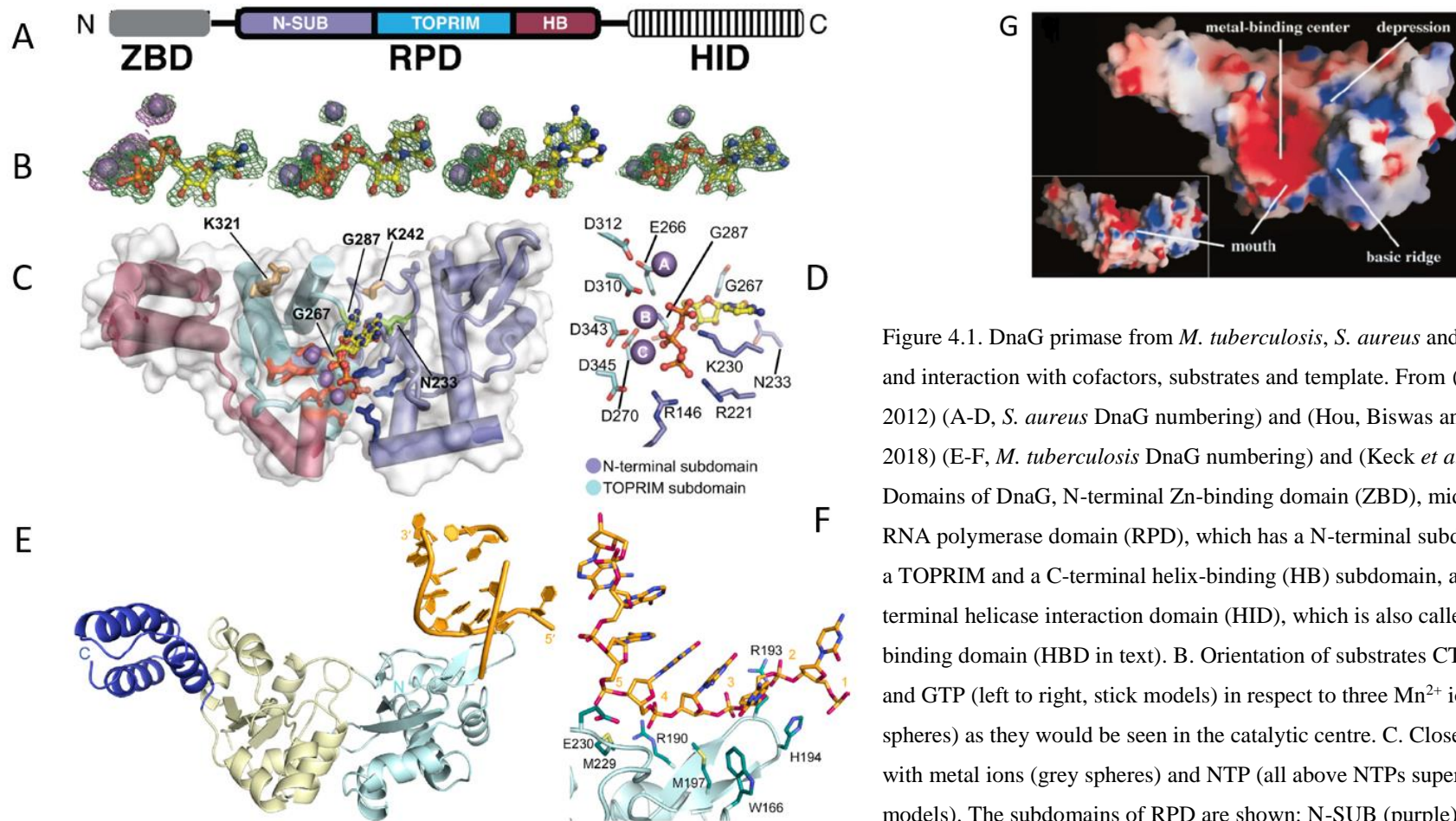


Figure 4.1. DnaG primase from *M. tuberculosis*, *S. aureus* and *E. coli* structures and interaction with cofactors, substrates and template. From (Rymer *et al.*, 2012) (A-D, *S. aureus* DnaG numbering) and (Hou, Biswas and Tsodikov, 2018) (E-F, *M. tuberculosis* DnaG numbering) and (Keck *et al.*, 2000) (G). A. Domains of DnaG, N-terminal Zn-binding domain (ZBD), middle section is RNA polymerase domain (RPD), which has a N-terminal subdomain (N-SUB), a TOPRIM and a C-terminal helix-binding (HB) subdomain, and the C-terminal helicase interaction domain (HID), which is also called helicase binding domain (HBD in text). B. Orientation of substrates CTP, UTP, ATP and GTP (left to right, stick models) in respect to three Mn^{2+} ions (grey spheres) as they would be seen in the catalytic centre. C. Close-up of RBD with metal ions (grey spheres) and NTP (all above NTPs superimposed, stick models). The subdomains of RPD are shown: N-SUB (purple), TOPRIM

(bright blue) and HB (pink). Amino acid restudies of basic ridge are dark blue sticks, metal binding residues are red sticks. Indicated numbered residues are likely involved in NTP binding. D. close-up of catalytic centre residues interacting with three metal ions (spheres A, B and C) and CTP (orange/yellow). E. RBD of *M. tuberculosis* interacting with dsDNA (yellow) via HB (light blue). Also TOPRIM (off-white) and N-SUB domains (dark blue) are shown. F. Close-up of residues of HB (dark teal) interacting with DNA (yellow). G. Naming of subdomains within the RPD of *E. coli* DnaG. Note relative positions of metal binding centre and basic ridge.

DnaG also interacts with SSB, a connection which is broken after primer is completed, resulting in dissociation of DnaG (Marceau *et al.*, 2011; Naue *et al.*, 2013). The clamp loading complex pentamer is held together by proteins ψ (*holD*) and χ (*holC*), which bind SSB during elongation stage (Marceau *et al.*, 2011).

Bacterial replication is initiated frequently on lagging strand, starting with primer synthesis. Likely, one Pol III core is responsible for continuous synthesis at leading strand and two are involved in lagging strand synthesis. According to the “trombone model” (Yoshikawa, 1970), the lagging strand successively loops out in the replication fork, and is bound by SSB for protection, to be replicated by Pol III into 5’-direction, while leading strand is replicated continuously. In a “polymerase switching” mechanisms, the lagging strand is synthesised by one Pol III and a “reserve” Pol III which is available to replace the active Pol III in cases of replication blocks. This was observed with EM in phage T7 replication (Hamdan *et al.*, 2007). Alternatively, both Pol III are active on the lagging strand simultaneously, at start and end of successive Okazaki fragments. This was observed by EM for phage T4 replication (Yang *et al.*, 2005).

In bacteria, Okazaki fragment maturation and replication repair are directed over β sliding clamps which can connect to up to 3 different polymerases and factors simultaneously. Okazaki fragments are processed by Pol I and DNA ligase I (Lopez de Saro & O’Donnell 2001). Pol I synthesises DNA starting at Okazaki 3’-ends, and cleaves primer via 5’-nuclease activity (Xu, Grindley and Joyce, 2000). Ligase seals the DNA backbone. Repair DNA polymerases II (Pol II) (*polB*), Pol IV (*dinB*) and Pol V (*umuCD*) are recruited to template breaks (Fuchs & Fuji 2007). Mismatch repair factor MutS repairs mismatches that have escaped proofreading during strand synthesis. As all these factors interact with the β clamp in the replication fork, it is referred to as “intrareplicative repair” mechanism (Lovett 2007).

Termination of replication occurs at termination (*ter*) sites on the *E. coli* chromosome (Siddiqui, On and Diffley, 2013; Dewar and Walter, 2017). Two replication forks approach each other, causing positive supercoiling which is, to some extent, released by various topoisomerases (Hiasa and Marians, 1996; Levine, Hiasa and Marians, 1998; Postow *et al.*, 2001; Espeli *et al.*, 2003). As replisomes enter *ter*, they are stalled. It is uncertain whether one is stalled as the other approaches (Dimude *et al.*, 2016), or they encounter and collide (Duggin and Bell, 2009). It has been documented that replication forks can converge (Levine, Hiasa and Marians, 1998; Espeli *et al.*, 2003), which can lead to re-replication, which is prevented by protein Tus (Hiasa and Marians, 1994; Markovitz, 2005; Rudolph *et al.*, 2013). Replication

fork converging can also lead to 3'-flaps, which can be removed by Pol I (Markovitz, 2005). After disassembly of replisomes, a non-replicated stretch may remain between them which must be filled by an unknown mechanism that involves protein PriA (Hiasa and Mariani, 1994; Rudolph *et al.*, 2013; Wendel, Courcelle and Courcelle, 2014).

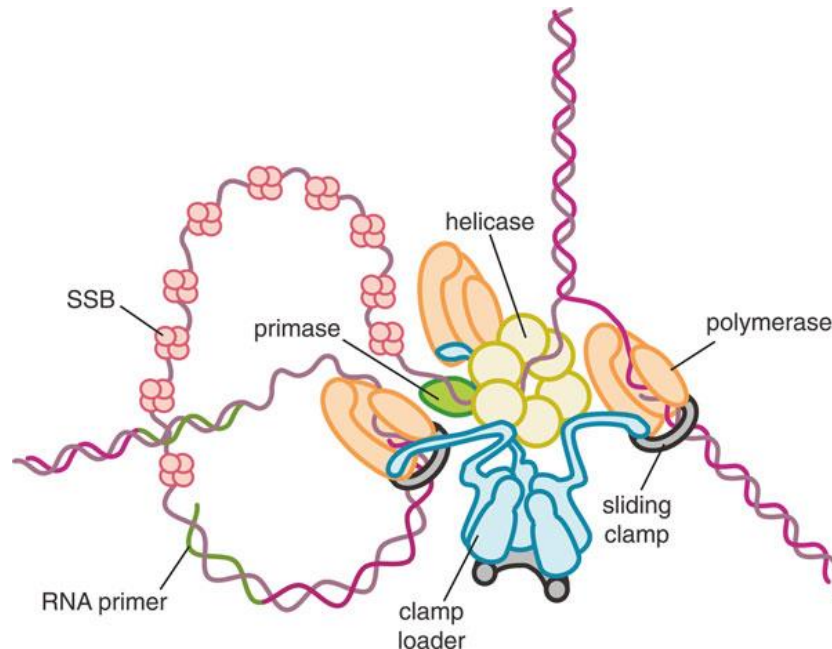


Figure 4.2 from (Kelch *et al.*, 2012). The bacterial replisome and trombone model. Clamp loader heteropentamer connects helicase homohexamers and three cores of polymerase and sliding clamps. Helicase progressively opens the replication fork, sliding clamps follow on dsDNA daughter strands, holding polymerases in place for DNA synthesis. SSB are necessary for looping of the lagging strand. Potentially, two polymerases can be engaged on the lagging strand, in this case two ssDNA loops would be observed. Primase associates to helicase transiently to create primer. Further factors and repair polymerases, as well as Okazaki fragment processing enzymes interact with the replisome in the replication fork.

4.2.3 Replication is regulated in stringent response via regulation of DnaG

Regulation of replication via DnaG is one pathway of stringent response. The stringent response reacts to nutrient starvation signalled by uncharged tRNAs. Protein RelA recognises non-aminoacylated tRNA bound to the ribosome, and phosphorylates GTP to form ppGpp or pppGpp (Winther, Roghanian and Gerdes, 2018). Another pathway is formation of (p)ppGpp from GTP by SpoT in response to carbon, iron and fatty acid starvation (Traxler *et al.*, 2008). Simultaneously, SpoT-mediated hydrolysis of (p)ppGpp is inhibited, which contributes to their accumulation in the cell. (P)ppGpp are called stress alarmones because they interact with various cellular targets to modulate rapid reduction of replication and biomass production (figure 4.3). (p)ppGpp were shown to bind DnaG at the catalytic centre, partially overlapping with bound NTP (see figure 4.3.B and C). While 5'-phosphates of (p)ppGpp are located like

those of GTP, 3'-phosphates change the orientation of base and ribose (Rymer *et al.*, 2012). Rymer *et al.* assume that alarmones disturb binding of both 3'-end RNA and incoming NTP competitively.

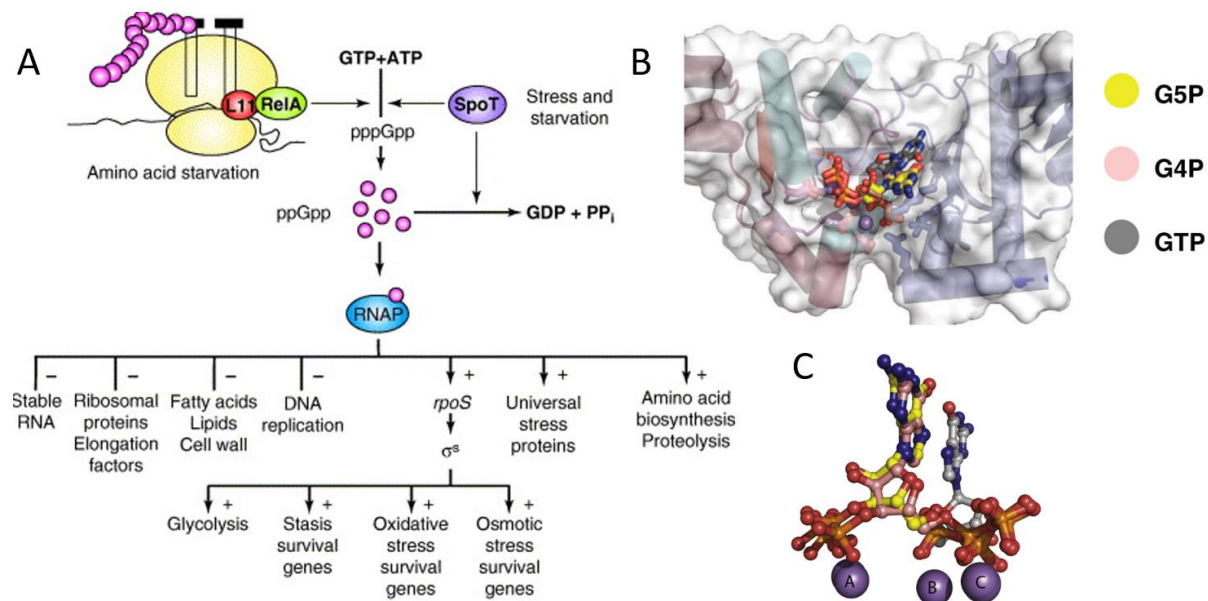


Figure 4.3 Stringent response and DnaG binding of alarmones. A. From (Magnusson, Farewell and Nyström, 2005). Stringent response is elicited by SpoT or RelA phosphorylation of GTP to form (p)ppGpp, in response to starvation signals. Probably, ppGpp binds to an allosteric site in RNAP holoenzyme and affects transcription of many genes and downstream processes. B. From (Rymer *et al.*, 2012). Besides targeting RNAP, (p)ppGpp also binds DnaG. Here, the alarmone binding site is almost identical to NTP binding site, thus (p)ppGpp competes with canonical substrates and inhibition depends on the relative concentrations of nucleotides and alarmones. C. From (Rymer *et al.*, 2012). Overlap of binding of GTP (grey), ppGpp (pink) and pppGpp (yellow) at DnaG metal ions (purple) in the catalytic site. Phosphates (orange/red) are coordinated almost identically on the metal ions, yet presence of 3'-phosphates slightly shifts the molecules above metal A such that ribose and nucleobase occupy a different space as compared to GTP which has a free 3'-OH.

4.2.4 Eukaryotic primer synthesis and removal differ from the prokaryotic system

Despite sequence and structural differences between bacterial, archaeal and eukaryotic replisomes, the mechanism of replication is quite similar (Kelch *et al.*, 2012). However, mechanisms of primer synthesis as well as structure of primase are quite dissimilar between bacteria and archaea/eukaryotes. Eukaryotic, archaeal and some viral primases are two subunit enzymes, generally consisting of a small catalytic and a large accessory subunit (Kuchta and Stengel, 2010).

Archaeal and eukaryotic primases are structurally similar despite sequence variation. Their mechanism of primer synthesis is analogous (Bell, 2019). Eukaryotic/archaeal primases have two (or three) subunits: catalytic subunit P49 (or P48 or PriS) and regulatory subunit P58

(also called PriL), accompanied by PriX in some archaea. In the following text I will refer to studies about archaeal subunits with using PriS, PriL and PriX, and eukaryotic primase will be referred to as P49P58

The P49P58 heterodimer complexes with DNA polymerase α (Pol α) to form pol-prim, an association that generates the most prominent difference between replication priming in eukaryotes and prokaryotes, the RNA-DNA primer (Boudet *et al.*, 2015). After 8-14 nt RNA primer is formed, primase is replaced by Pol α which attaches a 10-20 nt DNA extension to the RNA primer. Only then is DNA synthesis passed on to the processive DNA polymerases. While eukaryotic/archaeal primer are typically made of RNA, these primases readily incorporate dNTPs if abundantly available (Kuchta and Stengel, 2010; Ilic *et al.*, 2016) and thus can form full DNA primer, albeit at lower efficiency and with reduced primer length (Liu *et al.*, 2015).

P49 contains a Zn^{2+} -binding site which is important for protein structure, close to an active site which chelates cations via a cluster of conserved acidic amino acid residues (see figure 4.4 A and B). Further, a group of conserved basic residues function in NTP and DNA binding (Augustin, Huber and Kaiser, 2001; Ito *et al.*, 2003; Lipps *et al.*, 2004; Lao-Sirieix *et al.*, 2005; Yang, Lee and Nowotny, 2006; Vaithiyalingam *et al.*, 2014). In contrast to DnaG, P49 exhibits a two-metal-ion mechanism of catalysis (Kilkenny *et al.*, 2013). One divalent cation participates in NTP binding, and one in nucleotidyl transfer (Ito *et al.*, 2003). Conserved basic residues of P49 contact the triphosphate moiety of a bound nucleotide, where D 306, D111 hold Mg^{2+} , D109 simultaneously contacts both Mg^{2+} and all three phosphates, and R162 and H166 interact with γ and β phosphates, respectively. Further ribose-contacting residues have been identified (Vaithiyalingam *et al.*, 2014).

The primase large subunit contains a C-terminal iron-sulfur cluster (FeS-cluster or [4Fe4S]). Some archaea have besides PriL another PriX which also contains FeS-cluster and is structurally similar to PriL FeS-cluster C-terminal subdomain (Lao-Sirieix *et al.*, 2005; Klinge *et al.*, 2007; Weiner *et al.*, 2007; Pellegrini, 2012; Baranovskiy *et al.*, 2015; Kazlauskas *et al.*, 2018; Bell, 2019). P58 N-terminal half interacts with P49, while N-terminal P58 constitutes a pseudo-tandem repeat of the FeS-cluster subdomain. The catalytic subunit was found to predominantly execute primer elongation. Binding of the initiating nucleotide is facilitated at a conserved arginine in PriL/X or P58 (Kilkenny *et al.*, 2013; Baranovskiy *et al.*, 2016; Yan *et al.*, 2018). This binding site interacts with 5'-triphosphates of the initiator (Yan *et al.*, 2018). In the crystal structure of free primase, PriS active centre and PriL initiator

binding regions are distal, but upon initiation of primer synthesis configurational changes of the primase heterodimer bring these sites closer together (Baranovskiy *et al.*, 2015; O'Brien, Holt, *et al.*, 2018) (figure 4.4 B and C). Likely, the FeS-cluster of P49P58 in complex with substrate NTPs bind to anionic DNA. Electrostatic interaction between FeS-cluster in this composition with NTPs and DNA cause a conformational change that enhances primase activity. Thus eukaryotic replication initiation has been proposed to be regulated by redox signalling via DNA charge and FeS cluster (O'Brien, Holt, *et al.*, 2018). Figure 4.4 A-C show the orientation of primase subunits and domains in free conformation (B) and DNA/NTP-bound conformation (C).

Bacterial primer synthesis is initiated after DnaG first binds to DNA unspecifically and then finds its start site. In contrast to this, eukaryotic primer initiation begins with NTP binding, followed by template binding and first phosphodiester bond formation. Then, the catalytic centre is repositioned for the next cycle (Boudet *et al.*, 2015). The precise DNA recognition site has not been defined, but most eukaryotic primer bear 5'-purines, and primase seems to prefer to initiate on pyrimidine-rich templates (Suzuki *et al.*, 1993; O'Donnell, Langston and Stillman, 2013). The initiation site is likely determined by relative availability of NTPs rather than a specific recognition site (Sheaff and Kuchta, 1993; Kirk *et al.*, 1997). P58 binds to DNA template and does not move as P49 proceeds to elongate the primer along the template. P49 counting of primer (~10 nt) is dictated by the maximum stretch between P58 template binding domain and P49 catalytic site (Xie, 2011) (see figure 4.4 D). The primer is then handed over to Pol α via a yet unknown mechanism (Boudet *et al.*, 2015).

At least two pathways are known for eukaryotic primer removal, the "short flap" and "long flap" pathways. In the short flap pathway, which is the main pathway, processive DNA polymerase δ displaces RNA/DNA primer forming a "flap" between two dsDNA Okazaki fragments. Pol δ -associated protein flap endonuclease FEN-1 binds the base of a flap of about 10 nt, treads the free 5'-end through itself, and cleaves nucleic acid at the base of the flap endonucleolytically. This process is repeated a few times, eventually the RNA primer is removed and DNA ligase 1 seals the nick. Often, a part of an Okazaki fragment is removed which has been proposed to improve overall replication fidelity (Bambara, Murante and Henricksen, 1997; Lieber, 1997; Gloor, Balakrishnan and Bambara, 2010; Balakrishnan and Bambara, 2013). In the alternative long flap pathway, FEN-1 is not as readily available to

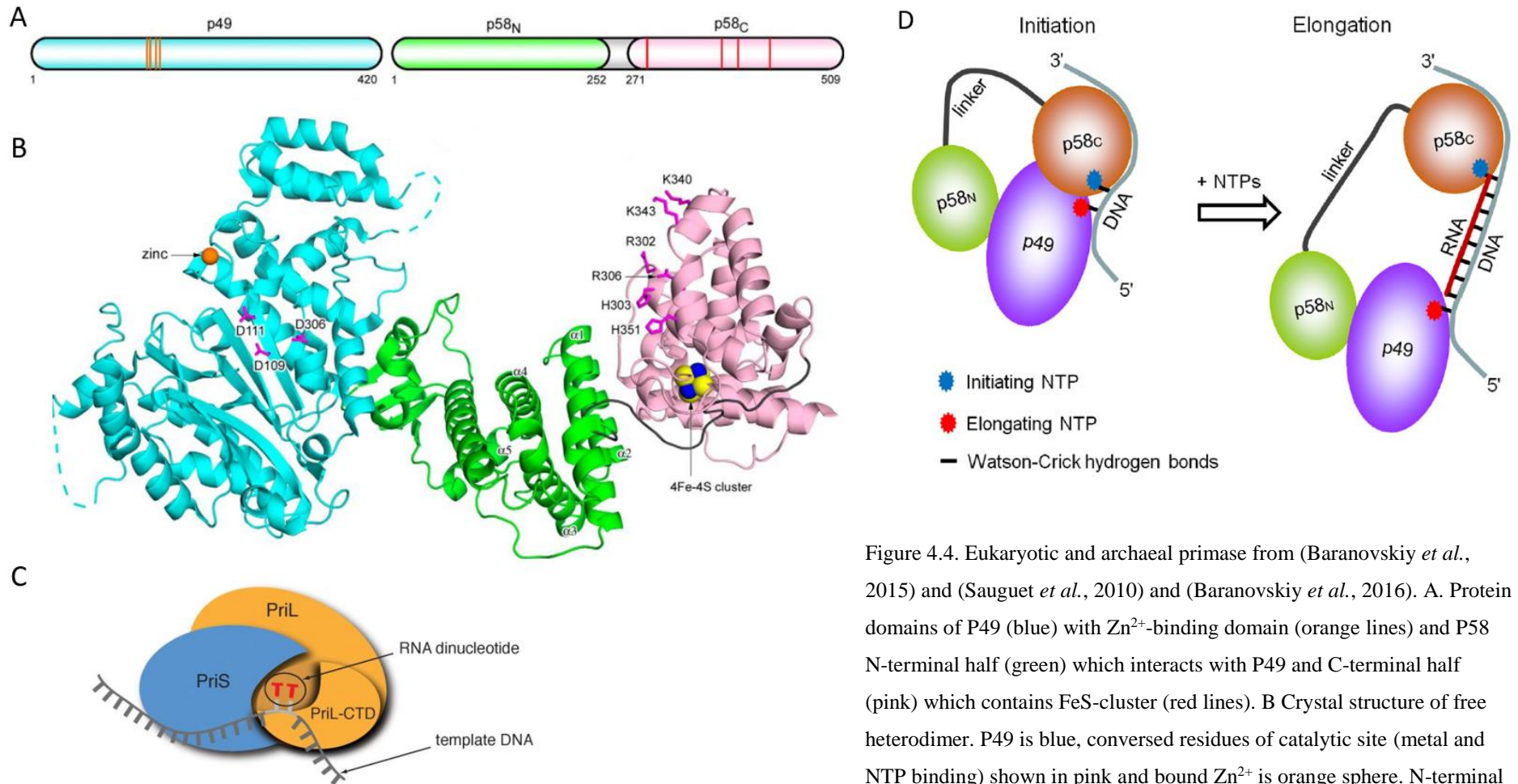


Figure 4.4. Eukaryotic and archaeal primase from (Baranovskiy *et al.*, 2015) and (Sauguet *et al.*, 2010) and (Baranovskiy *et al.*, 2016). A. Protein domains of P49 (blue) with Zn²⁺-binding domain (orange lines) and P58 N-terminal half (green) which interacts with P49 and C-terminal half (pink) which contains FeS-cluster (red lines). B Crystal structure of free heterodimer. P49 is blue, conversed residues of catalytic site (metal and NTP binding) shown in pink and bound Zn²⁺ is orange sphere. N-terminal domain of P58 (green), C-terminal domain (light pink) lie between P49 catalytic centre and FeS-cluster (yellow and blue spheres) and NTP- and DNA binding residues (pink). C. Orientation of PriS and PriL if bound to first two nucleotides and ssDNA. The nucleic acid binding sites, which are distal in free conformation, are now neighbouring. D. Primer synthesis starts as P49P58 dimer binds template DNA and two nucleotides. First, the second nucleotide is bound, then the initiating nucleotide (Boudet *et al.*, 2015). P49 catalyses phosphodiesterbond formation and proceeds along the ssDNA template until the maximal distance to the stationary P58 forces termination of primer at ca. 10 nt.

cleave short flaps repetitively, thus the flap becomes long enough for replication protein A (RPA) to bind. RPA is the eukaryotic equivalent to SSB, it binds ssDNA loops during replication to prevent reannealing, but is additionally involved in recruitment of further replication factors. FEN-1 cannot bind to RPA-DNA, but Dna2 protein can bind the base of the flap, remove RPA, and thread the free flap through its arch similarly to FEN-1. While threading, it cleaves the flap every 5-6 nt. Yet, it cannot cleave the base of the flap. Here, FEN-1 or its replacement Exo1 performs the final cleavage at the flap base (Bae *et al.*, 2001; Rossi and Bambara, 2006; Stewart *et al.*, 2008; Kang, Lee and Seo, 2010; Budd *et al.*, 2011).

Further, an RNase H-mediated pathway of primer removal exists in eukaryotes. RNase H1 removes primer of at least 4 nt length, RNase H2 removes primer of at least 2 nt length. Both cleave endonucleolytically between two ribonucleotides, but cannot cleave the RNA-DNA junction. This has to be conferred by another nuclease, such as FEN-1 (Turchi *et al.*, 1994; Murante, Henricksen and Bambara, 1998; Cerritelli and Crouch, 2009).

4.3 Aims

Here, we wanted to test whether bacterial primase DnaG and human primase catalytic subunit P49 initiate RNA synthesis *in vitro* with nucleotide-containing metabolites, and if so, which DNA and protein structural determinants are involved in this process. We have previously seen that different RNA polymerases can initiate transcription with abundant cellular cofactors. These transcriptases include structurally unrelated multi-subunit and single-subunit enzymes. Additionally, POLMRT readily produces NCIN-capped transcripts which can function as primer in mitochondrial replication.

Further, we intended to investigate the potential downstream effect of utilisation of metabolic cofactors in primer synthesis by DnaG on bacterial replication. Also, we wondered if nutrient availability in the cell, as signalled via stringent response alarmones (p)ppGpp, affects generation of non-canonical primer by DnaG *in vitro*.

4.4 Results

4.4.1 *DnaG initiates primer synthesis with metabolic cofactors NAD(H), FAD and DP-CoA at varying efficiency*

In order to test whether primer RNA can be synthesised by DnaG using adenine-containing cofactors, we designed an ssDNA template with the DnaG recognition site 5'-CTG (figure 4.5 A). Interestingly, we observed that DnaG initiates primer with NADH and FAD efficiently. K_M for ATP, NADH and FAD were 47, 109 and 390 μ M, respectively. Incorporation of NAD⁺ and DP-CoA was less efficient. Figure 4.5 B shows the different migration for primer initiated with cofactors. Figure 4.3.1 C further shows that metabolite initiation does not lead to the formation of abortive products. This finding is an analogy to NCIN-associated promoter escape in transcription (chapter I).

4.4.2 *DnaB improves specificity of primer initiation with canonical and non-canonical substrates*

In the cell, DnaG associates with helicase DnaB and SSB. It was known that DnaB strongly stimulates DnaG activity (Lu *et al.*, 1996; Johnson, Bhattacharyya and Griep, 2000). We tested whether presence of DnaB would affect the efficiency of DnaG NCIS incorporation. We could observe two effects of DnaB: i) DnaB enhances DnaG activity, ii) DnaB does not alter the relative efficiency of incorporation of substrates (ATP>NADH>FAD). Furthermore, in presence of DnaB “overlong primer” are not produced. Overlong primer are side-products of *in vitro* primer synthesis likely caused by template-independent extension of primer by DnaG (Bhattacharyya and Griep, 2000).

Primase was also reported to have a “counting” ability that ensures primer are usually 10-12 nt in length. This can be observed in form of a secondary band (ATP-10 in figure 4.6).

4.4.3 *NudC can hydrolyse 5'-NADH on primer*

In the case of NCIN-capped transcripts, NudC is a useful tool for verification of the NAD(H)-cap. Thus, we tested whether NudC would also cleave NADH-primer. Theoretically, the hybridisation of RNA primer to DNA might hinder NudC binding to its substrate. Here we show that if primase is washed off the hybrid (with 1M salt), NADH-primer is susceptible to NudC decapping (figure 4.7).

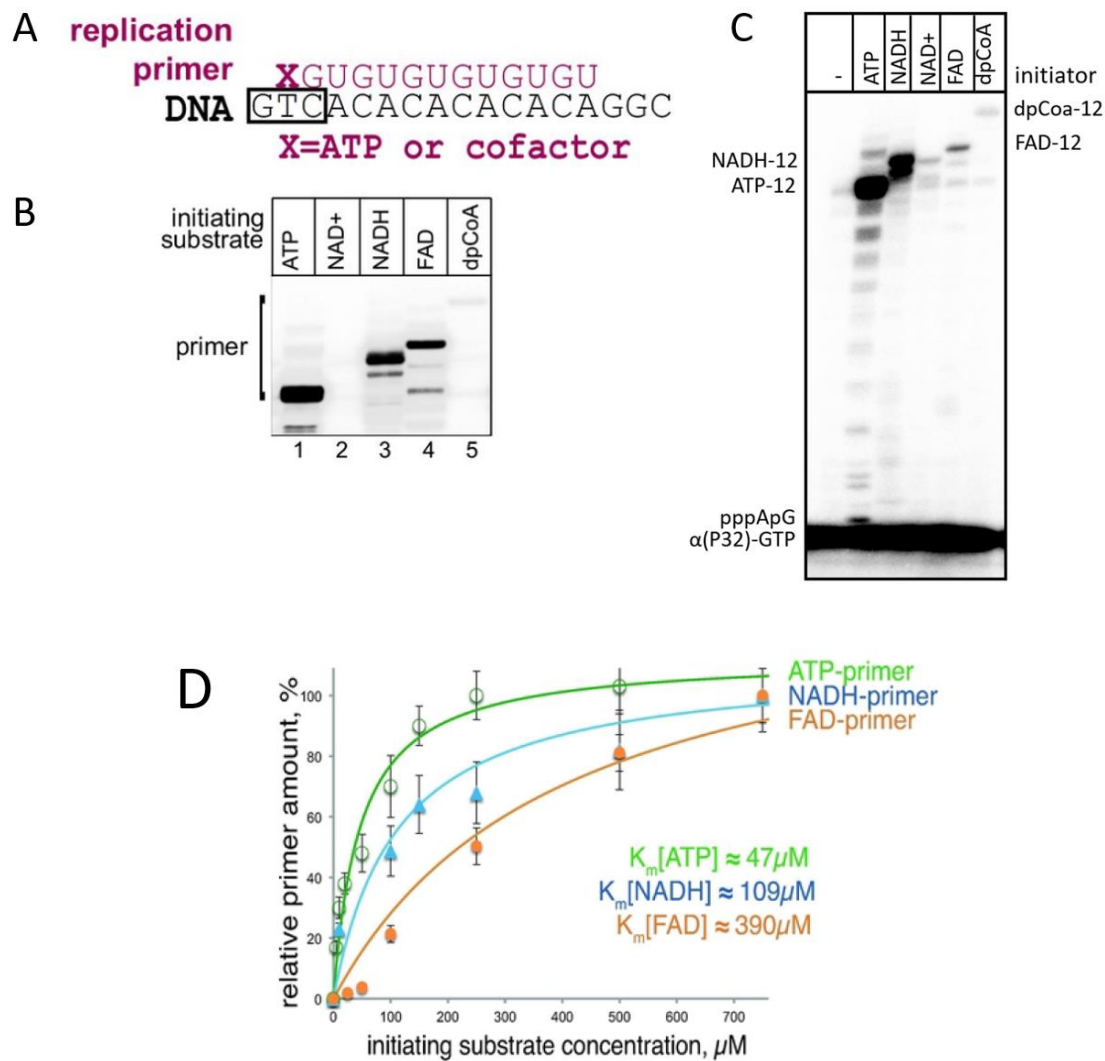


Figure 4.5 DnaG utilises adenine-containing cofactors of metabolism to initiate primer synthesis on ssDNA. A. Template sequence including recognition site 5'-dCTG (purple). The resulting RNA primer starts with 5' ATP or ADP-containing cofactor. +1A and +2G are shown in purple. The assay contained radiolabelled $\alpha(P^{32})$ -GTP and UTP for elongation to yield an up to 13 nt long primer. B. Denaturing polyacrylamide gel electrophoresis (23%) allows for resolution of primer initiated by different cofactors. We can observe that full length primer (12 nt + initiator) is produced upon initiation with ATP (lane 1), NADH (lane 3) and FAD (lane 4) while NAD⁺ and DP-CoA (lanes 2 and 5) make poor substrates for DnaG. Additional bands are result of either utilisation of alternative start site (e.g. initiation at +2) or shorter than full length primer (e.g. NADH-11), which can be attributed to primer counting by DnaG. C. Migration of cofactor- or canonical primer and abortive products of primase reaction. Similar to RNAP, NCIS does not promote release of abortive products by DnaG. D. Kinetics for incorporation of ATP (green), NADH (blue), or FAD (orange) into 5'-primer by DnaG. Michaelis-Menten kinetics were calculated for an average from three independent experiments Dots and error bars are mean values \pm standard deviation of three independent experiments.

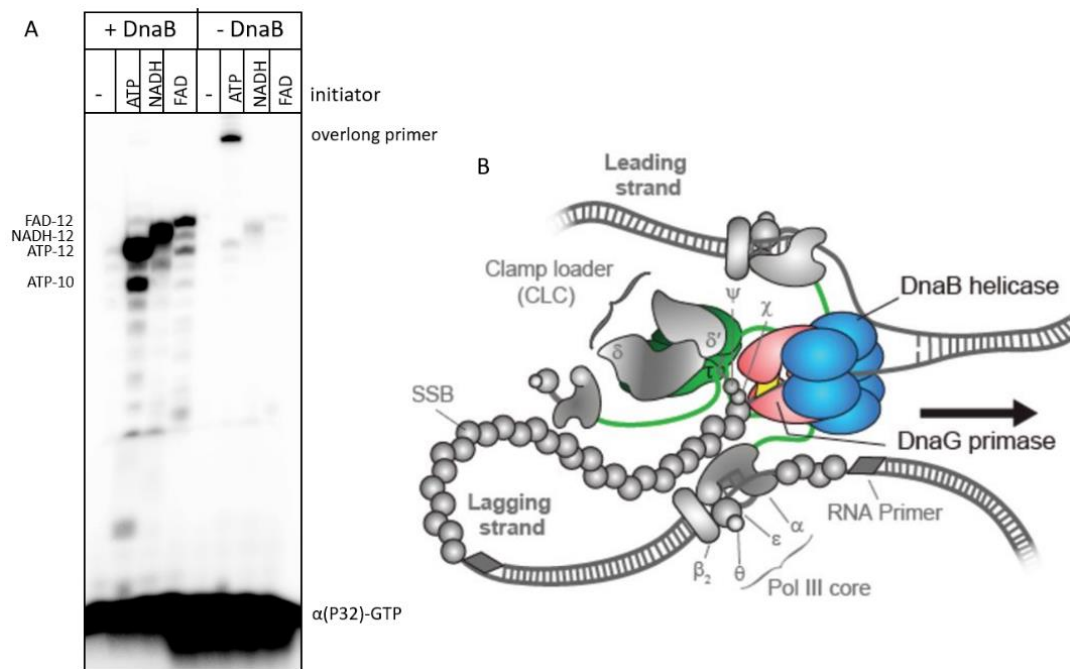


Figure 4.6 DnaB stimulates DnaG activity. A. Addition of DnaB helicase (left hand side of gel) leads to much stronger signal of radiolabelled primer as compared to absence of DnaB (right hand side). Also, production of overlong primer is reduced in presence of DnaB. B. From (Monachino *et al.*, 2018). DnaB helicase (blue) undergoes temporary interaction with DnaG (red) during primer synthesis. Many primer are generated on the lagging strand.

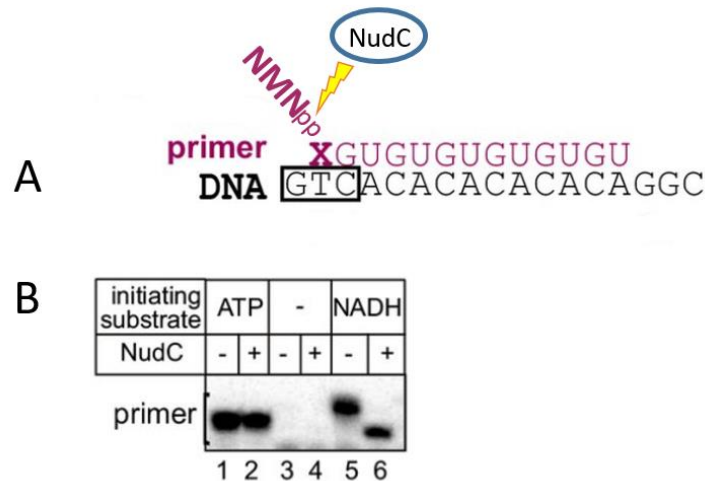


Figure 4.7. NudC can be used to verify 5'-NADH-primer. A. Schematic of NudC reaction on NADH-primer. NudC cleaves the phosphodiester bond between nicotinamide mononucleotide (NMN) and adenine mononucleotide (AMP). We cannot say if primer remains fully or partially annealed to the DNA template. B. After primer synthesis with ATP or NADH initiator, DnaG was washed off the template by addition of 1 M NaCl, and Nickle-NTA agarose beads (via His-tag on DnaG). The reaction was filtered through MicroBio Spin 6 column for buffer exchange, then 10 pmol NudC were added in 20 mM Tris-Hcl pH 8, 50 mM KCl and 10 mM MgCl₂. In absence of DnaG, primer is sensitive to NudC cleavage.

4.4.4 Cofactor modified 5'-primer RNA does not interact with DnaG recognition sequence -1 position on DNA template

For bacterial transcription initiation with non-canonical substrates by RNAP Bird *et al.* found that -1 position of promoter can modulate NCIN-capping efficiency (Bird *et al.*, 2016), although we could not confirm this (Julius and Yuzenkova, 2017). In the DnaG recognition site 5'-CTG, G functions as -1 position. As the non-adenosine moiety of cofactors could potentially interact with -1 base, we tested whether variation of this position (see figure 4.8 A) would affect cofactor incorporation efficiency. Reciprocally, if cofactors would lead to preference of alternative recognition sites, this might affect replication initiation frequency on the chromosome. We found that DnaG favoured recognition site does not change upon cofactor initiation (figure 4.8 B). In general, DnaG prefers -1 purines, with -1A being most efficient alternative to -1G. Therefore we conclude that X-moiety of cofactor does not interact with -1 template.

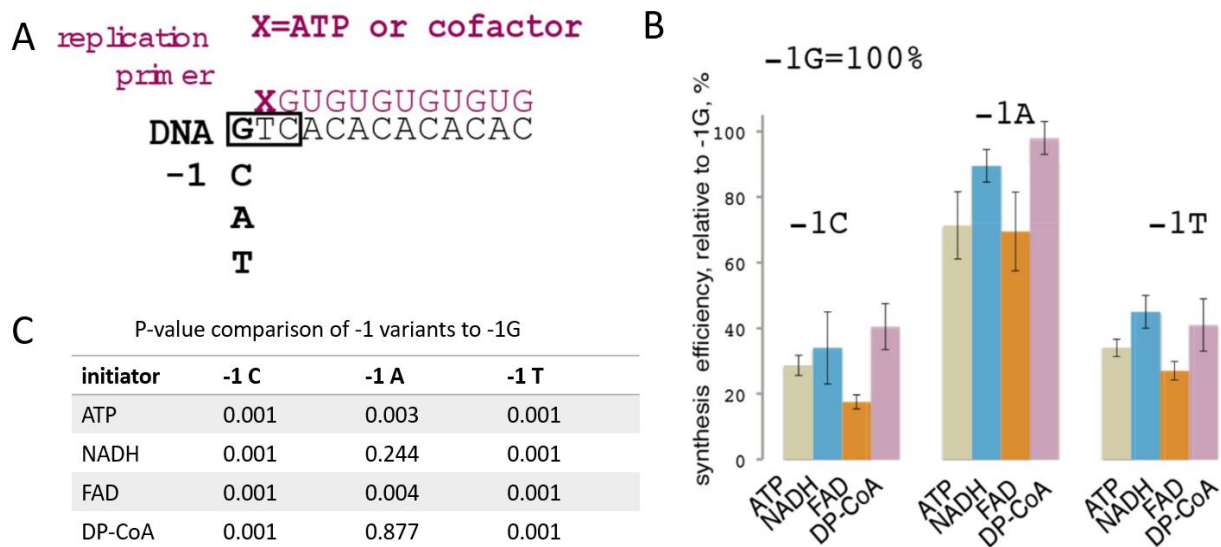


Figure 4.8. -1 template base affects initiation efficiency with canonical and non-canonical substrates, but does not affect relative efficiency of cofactor incorporation as compared to ATP. A. We varied template -1 position from G to C, A, and T on the DNA template. B. If compared to initiation at -1G template, DnaG can synthesise primer on -1 variants at varying efficiency. -1A is preferred alternative, while -1C and -1T are equally unfavourable. All NCIS are incorporated at relatively similar rate when compared to ATP. Thus, an interaction of NCIS and -1 is not likely. ATP: olive green, NADH: blue, FAD: orange, DP-CoA: purple. Standard variation is depicted from 3 independent experiments. C. P-values for experiments depicted in B, calculated with Tukey algorithm.

4.4.5 DnaG basic ridge residues are involved in binding of non-canonical initiating substrate

To assess potential points of contact between NCIS and DnaG protein we mutated residues of “basic ridge” and catalytic site which had previously been documented to be involved in initiation (Rymer *et al.*, 2012). Mutants K229A, K241A and Y230A (basic ridge mutants) (figure 4.9 A) were overall less efficient in primer synthesis than wildtype DnaG. Additionally, their relative efficiency of incorporation of NADH and FAD compared to ATP was reduced as compared to wildtype DnaG (figure 9 B). Mutant D309A (metal binding mutant) was not impaired in NCIS-primer formation (figure 4.9 B), albeit overall lower efficiency was observed. We hypothesise that basic ridge residues are involved in binding of substrate NTPs and potentially NMN and FMN moieties of initiating cofactors.

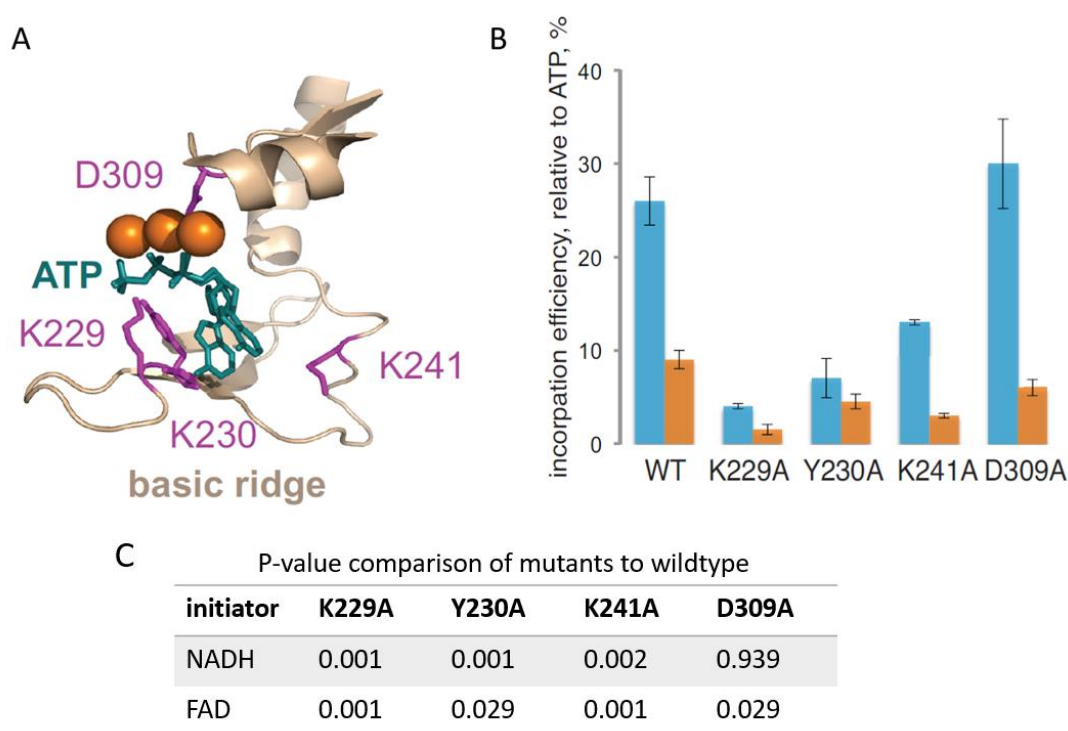


Figure 4.9. DnaG residues of basic ridge are likely involved in NCIS binding. A. Adapted from crystal structure (Rymer *et al.*, 2012; Hou, Biswas and Tsodikov, 2018). Wildtype DnaG residues that were mutated (pink) in spatial relation to catalytic metal ions (orange) and bound nucleotide (green). All residues were substituted to alanine in mutants. B. Relative incorporation efficiency of cofactors compared to ATP. wildtype DnaG (wt) incorporates NADH (blue) at about 25% efficiency as compared to ATP, and FAD (orange) only about 10%. Mutants K229A and Y230A were strongly inhibited in non-canonical initiation, K241A was also less efficient in relative incorporation. In contrast, D309A was able to incorporate both cofactors similarly to wt protein. Bars and error bars are mean values \pm standard deviation of three independent experiments. C. P-values of the comparison of efficiency of incorporation of NADH or FAD by mutants to incorporation efficiency by wildtype DnaG (three independent experiments). P-values were generated using Tukey algorithm.

4.4.6 Stringent response alarmone ppGpp inhibits NCIS-primer synthesis

Stringent response is a multifaceted mechanism that allows for selective reduction of metabolism and cell growth in response to amino acid starvation (Traxler *et al.*, 2008; Winther, Roghanian and Gerdes, 2018). When ribosome-associated protein RelA senses non-aminoacylated tRNA, it converts GTP to “alarmones” ppGpp or pppGpp. These alarmones are known to bind RNAP holoenzyme to downregulate production of ribosomes and many other gene products. (P)ppGpp was shown to inhibit *E. coli* primase *in vitro* (Maciąg *et al.*, 2010). Alarmone inhibition of replication *in vivo* is still subject of debate (Potrykus *et al.*, 2011; Maciąg-Dorszyńska, Szalewska-Pałasz and Węgrzyn, 2013). (P)ppGpp was shown to bind in the DnaG active site (Rymer *et al.*, 2012), thus competition with canonical substrates was suggested as mechanism of inhibition.

We tested whether ppGpp affects NCIS incorporation. PpGpp is always present in the cell at low level. Different studies report exponential phase levels of (p)ppGpp between few μM (Maciąg *et al.*, 2010) and on average 40 μM (Varik *et al.*, 2017). (P)ppGpp levels then rise to several hundred μM during stringent response (Maciąg *et al.*, 2010; Varik *et al.*, 2017) with a maximal accumulation up to 900 μM (Traxler *et al.*, 2008).. Average stationary phase concentration of ppGpp is approximately 150 μM (Varik *et al.*, 2017). Thus we tested whether metabolites compete with ppGpp to allow for primer initiation. We saw that NCIS incorporation is inhibited at distinctively lower ppGpp concentration than ATP (figure 4.10). We therefore propose that NCIS-primer formation is downregulated earlier during nutrient stress. An estimated stationary phase concentration of ppGpp around 150 μM would reduce incorporation of FAD by >70% and NADH by >40%.

4.4.7 5'-cofactors differentially affect primer removal by DNA polymerase I

Maturation of Okazaki fragments is required for formation of the lagging DNA strand. For this, primer needs to be removed and replaced by DNA, before DNA backbone nicks can be ligated. Pol I fulfils two functions, it degrades RNA primer by 5'-3'-exonuclease activity, and polymerises DNA in 5'-3' direction to fill the gap between two Okazaki fragments (Xu *et al.*, 1997). A non-base paired 5'-end of NCIS might impair or promote Pol I exonucleolytic function. Here, we document that Pol I can remove NCIS-primer, but at varying efficiency. Presence of 5'-NADH strongly enhances primer removal by Pol I, while 5'-FAD inhibits it (figure 4.11).

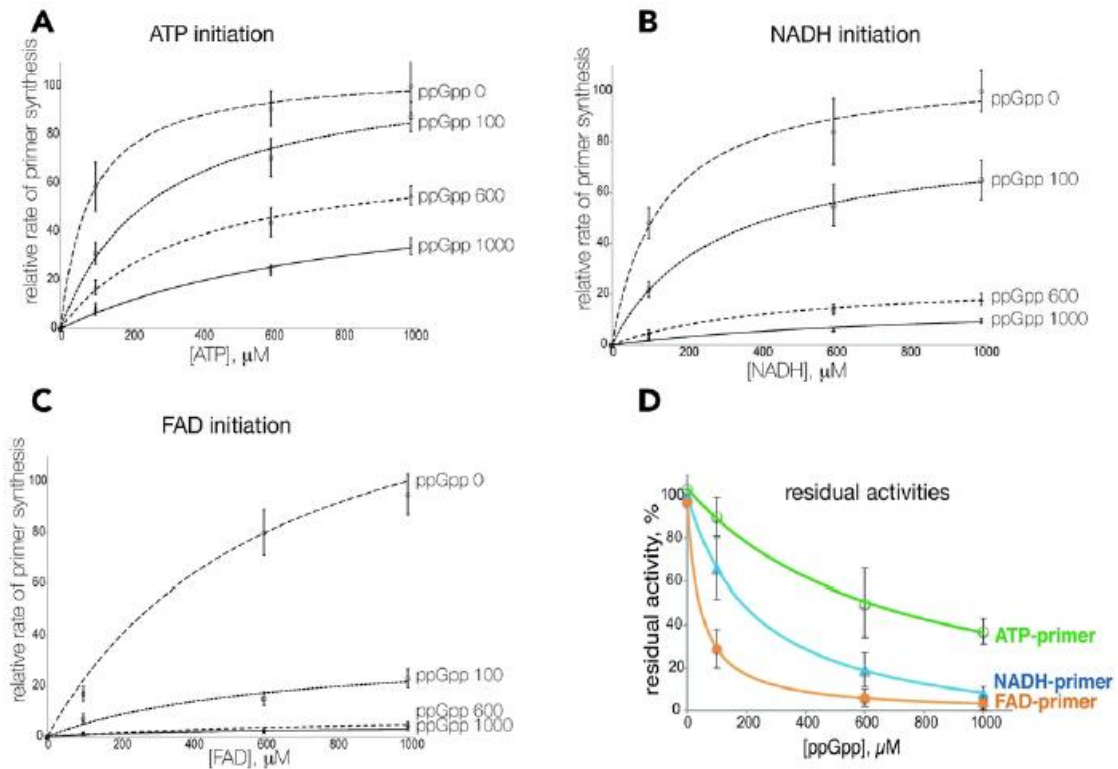


Figure 4.10 Inhibition of NCIN-primer synthesis by ppGpp. A-C. Measurement of primer synthesis at increasing concentrations of substrates ATP, NADH or FAD in presence of 100, 600, and 1000 μM ppGpp, or control (ppGpp 0). D. Increasing concentrations of ppGpp compete with other substrates (all 1000 μM). Primer synthesis with the canonical substrate ATP (green), was reduced up to 60%. Inhibition of non-canonical initiation is inhibited at lower concentrations of ppGpp. PpGpp competes with both FAD (orange) and NADH (blue). Dots and error bars are mean values \pm standard deviation of three independent experiments.

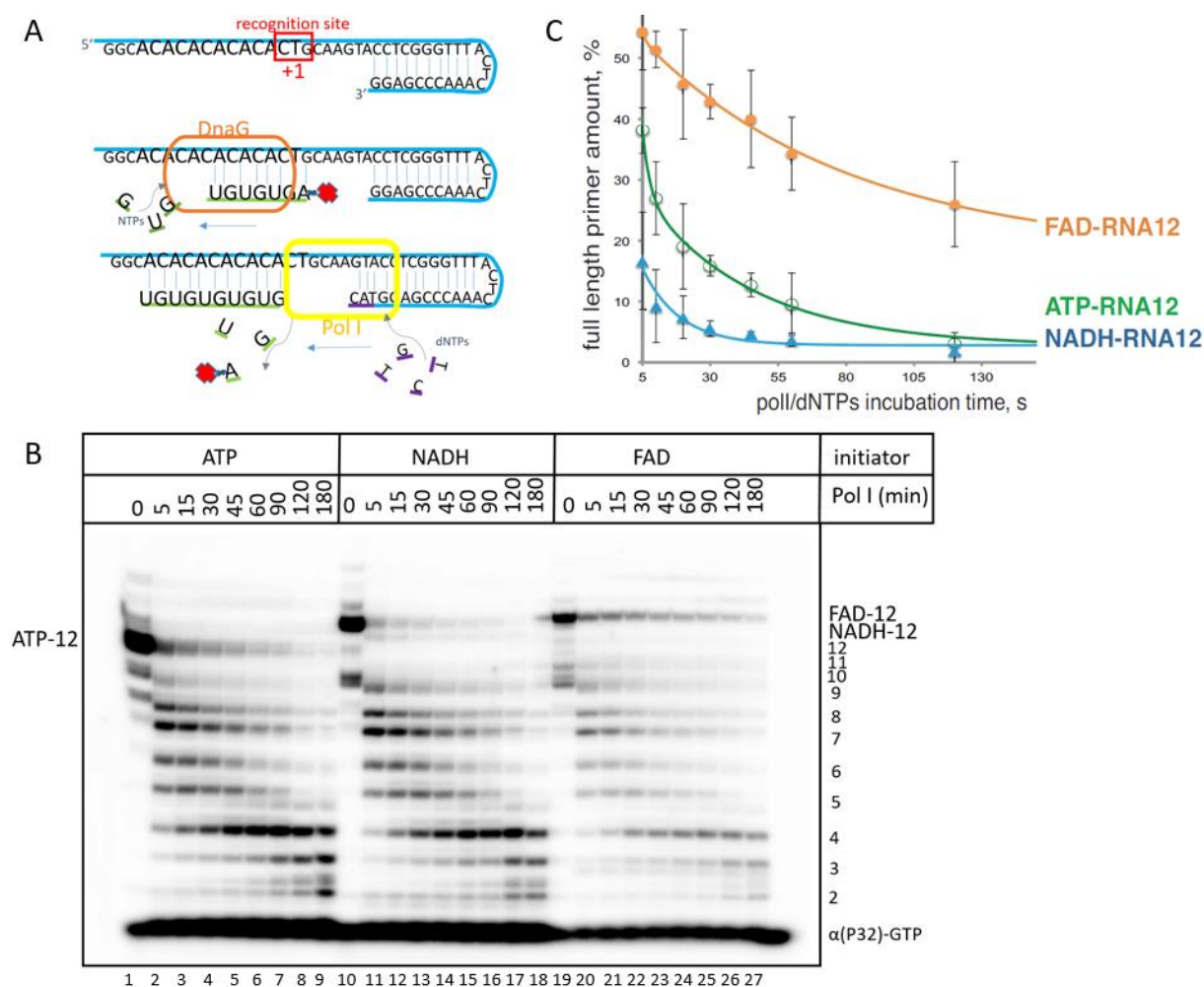


Figure 4.11 Primer removal by Pol I is affected by 5'-cofactors. **A.** Template and scheme of the reaction. We designed a hairpin template to present a free 3'-end upstream of the primase recognition site. After primer synthesis initiated by ATP, NADH or FAD (red X-A), DnaG was removed with 1M salt and Ni^{2+} -NTA agarose beads, and the reaction was filtered through Micro-Bio Spin 6 gel filtration columns, then Pol I was added to the hybrid/template. Pol I binds the 3'-end of the hairpin and polymerises dNTPs. As it approaches the primer, the primer is displaced and broken up by 5'-nuclease activity. **B.** Denaturing polyacrylamide gel electrophoresis of a time course of Pol I reaction. The full length primer (ATP-12, NADH-12, or FAD-12) faded rapidly after addition of Pol I. Bands appear underneath the full primer which are breakdown products of Pol I 5'-3'-exonuclease activity. **C.** Graph depicting the degradation of full length primer by Pol I over a time course of two minutes, in presence of $2.5 \mu\text{M}$ dNTPs. NADH-initiated primer is hydrolysed more rapidly than ATP-initiated primer. FAD-primer cleavage is strongly delayed. Dots and error bars are mean values \pm standard deviation of three independent experiments.

4.4.8 5'-cofactors do not affect primer removal by RNase H

Primer are short lived products which are degraded rapidly after they have fulfilled their purpose. In *E. coli* this is achieved by one of at least two different pathways, one of which is RNase H digestion. RNase H is an endoribonuclease with specificity for DNA-RNA hybrids (Nowotny *et al.*, 2005). Our assays show that presence of 5'-NCIS on primer does not interfere with RNase H-mediated primer removal, although longer products seem to arise as RNase H does not cleave phosphate bond within the cofactor (see figure 4.12). Primer degradation by RNase H yields oligoribonucleotides of varying lengths, some of which are susceptible to NudC cleavage. We observed that RNase H cleavage made NCIS-primer oligos more susceptible to NudC, while FAD-12 and DP-CoA-12 are insensitive to NudC (figure 4.12 B). It would be interesting to investigate in the future, whether any association between RNase H and decapping enzymes can be observed *in vivo*.

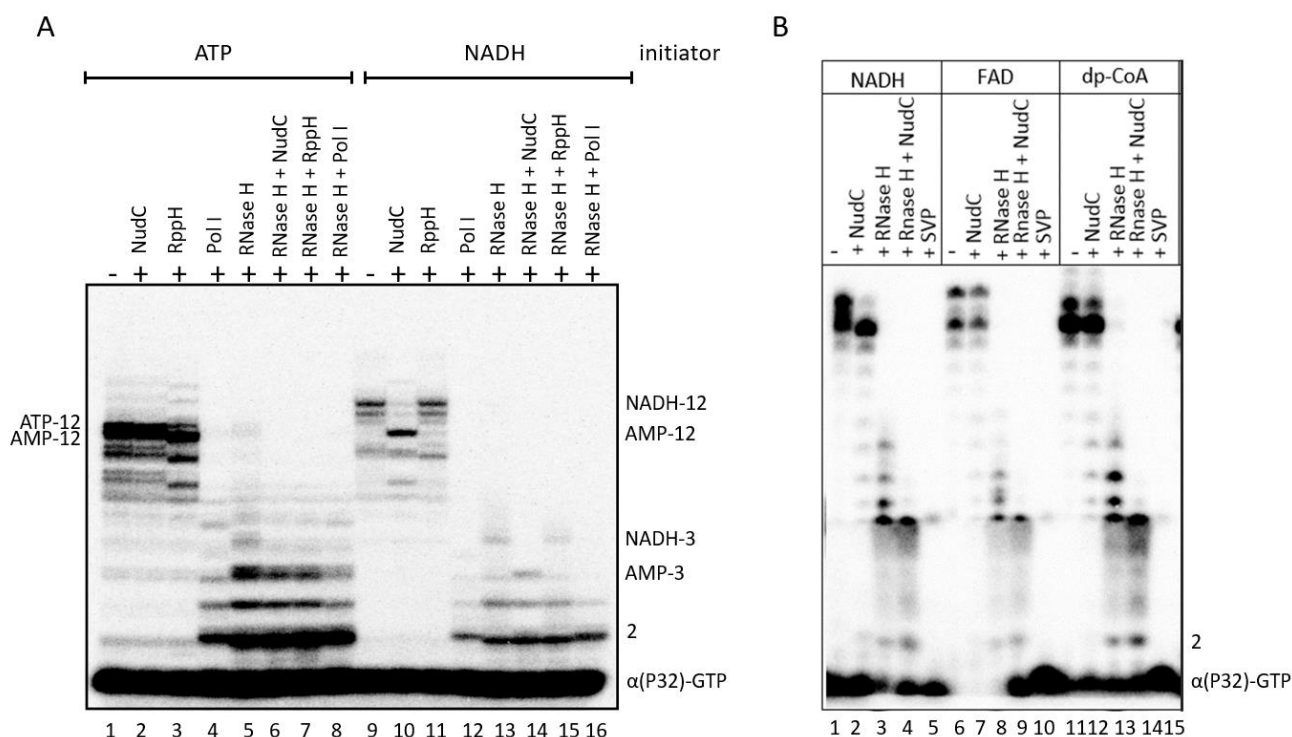


Figure 4.12. Cleavage of primer by RNase H and decapping enzymes or other nucleases. A. ATP-12 and NADH-12 primer were produced in a standard reaction. DnaG was washed off (lanes 1 and 9) and decapping enzymes or RNase H were added as indicated. ATP-12 is insensitive to NudC but RppH produces AMP-12 which runs slightly lower than ATP12. In contrast, NADH-12 is insensitive to RppH, but NudC creates AMP-12. B. As in A, NCIS-primer were produced by DnaG reaction and DnaG was removed. Addition of NudC alongside RNase H produced smaller oligos than RNase H alone. Snake Venom Phosphodiesterase (SVP) was added as control to ensure that DnaG was not bound to the hybrid. SVP cleaves all phosphodiester bonds, thus, removal of DnaG should result in mononucleotides.

4.4.9 The human primase catalytic subunit P49 can initiate primer synthesis with guanosine-analogues

Bacterial and eukaryotic replication function overall by a similar mechanism and have the same result, yet primer synthesis is vastly different. Eukaryotic primase is a two-subunit enzyme, with catalytic subunit P49 and regulatory subunit P58. Here we tested whether P49 would be able to initiate primer synthesis using cofactors. The exact recognition site on template DNA for P49 is unknown, though 5'-purines are most commonly found on eukaryotic primer. We thus designed a ssDNA template (figure 4.13 A) with *E. coli* DnaG recognition site 5'-CTG. However, P49 did not incorporate ATP. Only GpU and m7GTP were accepted by P49 as substrates for abortive initiation, albeit P49 did not elongate these abortive products (figure 4.13 B). As P58 subunit is involved in initiation and NTP binding, we are going to include it in future tests.

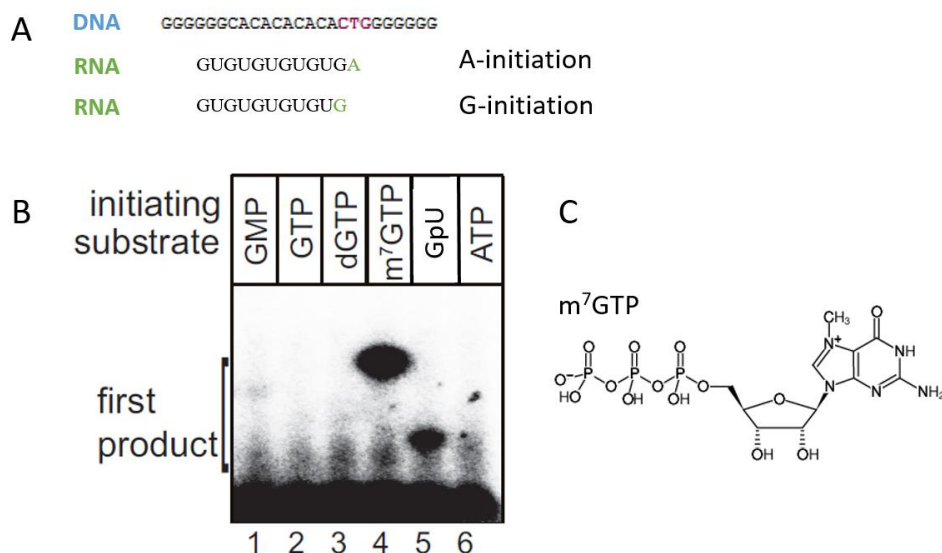


Figure 4.13. Incorporation of guanosine-containing small molecules by eukaryotic primase catalytic subunit P49. A. ssDNA template was designed similarly to an *E. coli* primase template. If P49 was to initiate synthesis incorporating 5'-A, a 12-mer could be synthesised from added GTP and UTP. P49 could also initiate with 5'-G, then up to 11 nt could be polymerised in a template-dependent manner. B. First trials with adenosine-containing initiators were unsuccessful, thus we tested incorporation of guanosine-containing molecules. We found that GpU (dephosphorylated dinucleotide) was polymerised to $\alpha(P^{32})$ -UTP (lane 5), and also a methylated GTP was accepted as substrate by P49 (lane 4). Synthesis of longer products from these initiators was not observed. C. Structural formula of methylated GTP.

Results presented in this chapter were submitted to the journal Nucleic Acids Research for publication in collaboration with Dr Yulia Yuzenkova. The manuscript is currently in review and was therefore not cited here. The manuscript can be viewed in the bibliography at the end of this thesis.

4.5 Discussion

4.5.1 *Bacterial primase DnaG readily initiates primer synthesis using nucleotide-containing cofactors of metabolism*

Previously, it was documented that dsDNA-dependent RNA Polymerases of prokaryotes, eukaryotes, mitochondria, and viruses can initiate template-dependent transcription using cellular metabolites which contain a nucleotide moiety (Chen *et al.*, 2009; Kowtoniuk *et al.*, 2009; Cahová *et al.*, 2015; Bird *et al.*, 2016; Jiao *et al.*, 2017; Julius and Yuzenkova, 2017; Walters *et al.*, 2017; Bird, Basu, Kuster, Ramachandran, Grudzien-Nogalska, Kiledjian, *et al.*, 2018; Julius, Riaz-Bradley and Yuzenkova, 2018; J. Wang *et al.*, 2019). Here we show that also ssDNA-dependent RNA polymerase DnaG is able to initiate primer synthesis using adenine-containing metabolites.

Bacterial primase DnaG preferentially initiates primer synthesis on triplet recognition site 5'-d(C₊₂T₊₁G₋₁) (Kitani *et al.*, 1985; Yoda and Okazaki, 1991). Thus, the 5'-end of primer constitutes an ATP, or here, an adenosine-containing metabolite. We observed that DnaG produces full length primer from initiation with NADH and FAD. Other NCISs, which were readily incorporated by EcRNAP in transcription, such as NAD⁺ and DP-CoA, proved to be poor substrates to DnaG (figure 4.5).

In bacterial primer synthesis, the first phosphodiester bond formation is the rate limiting step of, as the two first nucleotides are not stabilised by an adjoining RNA chain base paired to the template (Ilic *et al.*, 2018). Thus, an interaction of the nicotinamide or riboflavin of initiator with -1 template base could theoretically stabilise the initiation complex, as proposed for non-canonical transcription initiation by EcRNAP (Bird *et al.*, 2016; Vvedenskaya *et al.*, 2018). Similarly to our findings on NCIN-capping of *E. coli in vitro* transcripts (Julius and Yuzenkova, 2017), NCIS did not seem to interact with -1 template position in primer synthesis by DnaG (figure 4.8). Unfavourable interaction with substrate binding sites in the active centre may account for the reduced initiation efficiency for NCIS as compared to canonical substrate.

The catalytic centre of DnaG and the polymerisation mechanism are quite unique. While RNA transcriptases and DNA polymerases bind two divalent metal ions, bacterial primases bind three metal ions. (Aravind, Leipe and Koonin, 1998; Keck *et al.*, 2000; Kuchta and Stengel, 2010; Rymer *et al.*, 2012). The catalytic centre of DnaG shows high similarity to topoisomerases (which bind one metal cation for catalysis) (Rymer *et al.*, 2012). We tried to

determine how NCIS interact with DnaG conserved residues by mutation studies. We created mutants of the basic ridge, which were identified previously to be involved in substrate binding (Keck *et al.*, 2000; Rodina and Godson, 2006; Rymer *et al.*, 2012; Hou, Biswas and Tsodikov, 2018) or catalysis (Godson *et al.*, 2000; Rodina and Godson, 2006; Rymer *et al.*, 2012; Hou, Biswas and Tsodikov, 2018). In nucleotide triphosphates, the triphosphate group is attached to the 5'C of the ribose. In cofactors, in place of the triphosphate, a disphosphate linker connects ribose to an inverted functional moiety, such as nicotinamide mononucleoside (NMN) or flavin mononucleoside (FMN). The crystal structure of DnaG soaked with ATP (figure 4.9), shows phosphate groups reach across metal B and C while basic ridge residues around position 230 protrude towards them (Rymer *et al.*, 2012) (figure 4.1 and 4.9 A). Cofactor functional moieties are likely to affect the orientation of the substrate in the catalytic site (Hou, Biswas and Tsodikov, 2018). We substituted alanine for K229 and Y230 and observed that this substitution reduces DnaG ability to initiate with NADH and FAD drastically. Further, on the crystal structure it is visible that K241 in wt interacts with upstream template and potentially primer. Sun *et al.* documented that K241 mutants can initiate but not elongate primer (Sun, Schoneich and Godson, 1999). Therefore, we tested the effect of substitution of this residue to alanine on NCIS-primer initiation and elongation. K241A displayed reduced formation of NCIS-primer, but the effect was not as strong as mutations of residues K229 and Y230 (figure 4.9 B). Several acidic residues are involved in metal coordination in wt DnaG. D309 interacts with metals A and B as well as phosphate of NTP substrates. Mutation of this residue reduced overall activity of the enzyme but had no effect on NCIS incorporation. We conclude that basic ridge residues, which have been shown to coordinate triphosphate moiety of substrates, also accommodate NMN and FMN.

Thus far we have not been successful in verifying NCIS-primer synthesis *in vivo*. However, findings about NCIN-capping by transcriptases are encouraging. We believe that the *in vitro* ability of DnaG to incorporate NCIS efficiently in presence of DnaB, and high availability of these substrates *in vivo* (Bennett, Kimball and Gao, 2009), are in favour of our hypothesis.

4.5.2 NCIS incorporation by DnaG might have downstream effects on replication

The only function of primer is to provide a free 3'-OH for DNA polymerases to be able to start DNA synthesis. Synthesis of one primer requires about 0.2 sec (Kusakabe and Richardson, 1997; Frick, Kumar and Richardson, 1999) and per *E. coli* genome about 2000-4000 Okazaki fragments are generated with a size of approx. 1-2 kb (Ogawa and Okazaki, 1984; Swart and Griep, 1995; Lewis, Jergic and Dixon, 2016). Overall replication speed of *E.*

coli is around 1 kb/sec (Bailey, Wing and Steitz, 2006). As Pol III synthesises DNA at around 750 nt/sec, Okazaki fragment maturation is another rate limiting step of replication (Kusakabe and Richardson, 1997; Frick, Kumar and Richardson, 1999). Arguably, replication speed can be strongly affected by failure to remove primer RNA.

Okazaki fragment processing requires primer removal by *E. coli* Pol I (gene *polA*) and RNase H (gene *rnhA*) (Lewis, Jergic and Dixon, 2016). Deletion of either gene in *E. coli* is not lethal. RNase H seems to be dispensable for *E. coli* growth, whereas deletion of *polA* reduces bacterial growth (Konrad and Lehman, 1974). A double mutation ($\Delta rnhA \Delta polA$) is conditionally viable, also double mutants of $\Delta rnhA$ and thermo-sensitive Pol I have been produced in the lab (Konrad and Lehman, 1974; Ogawa and Okazaki, 1984; Kitani *et al.*, 1985; Kogoma, 1986). Pol I and RNase H cooperate in primer removal, and can also replace each other to some extent. In absence of both proteins, related enzymes, such as YpcP in *B. subtilis*, or other RNase H family genes (Fukushima *et al.*, 2007) can replace Pol I at a fitness cost.

Here, we tested whether Pol I activity would be affected by metabolites. Pol I was initially described as 5'-3'-exonuclease, yet reports are now conflicting (Xu *et al.*, 1997). Pol I was not found to release mononucleotides (Kelly *et al.*, 1969; Xu *et al.*, 1997), which contrasts our presented results (figure 4.11). Pol I is also described as a structure-specific endonuclease that recognises ss/ds junctions or displaced 5'-"flaps". Therefore the name was revised to 5'-nuclease (Xu, Grindley and Joyce, 2000). Pol I is involved both in DNA repair (nick translation) and primer removal (Lundquist and Olivera, 1982; Lyamichev, Brow and Dahlberg, 1993; Xu *et al.*, 1997). In primer removal, the polymerase domain of Pol I displaces the 5'-end of primer which "flaps out" and thus becomes sensitive to structure-specific endonuclease activity (Lundquist and Olivera, 1982; Xu *et al.*, 1997). In fact, 5'-nuclease domain of Pol I was shown to have structural similarity to eukaryotic 5'-flap endonuclease FEN-1, which removes primer in eukaryotic replication (Harrington and Lieber, 1994; Robins *et al.*, 1994). We have shown that cofactor non-adenine moiety does not base pair with -1 base on DNA (figure 4.8), thus we saw a possibility that cofactor initiation might aid "flapping out" of primer, which might support 5'-structure-specific nuclease activity of Pol I. Indeed, NADH primer removal by Pol I was more efficient as compared to ATP-primer (figure 4.11). However, FAD-primer removal by Pol I was found to be delayed. We have to assume that Pol I interacts with 5'-FAD unfavourably. The 5'-nuclease domain of Pol I contains an arched structure of two helices. One helix is clad with positively charged residues, the other with hydrophobic residues. The nuclease active site is at the bottom of the arch

where two metal ions are bound (Ceska *et al.*, 1996; Hosfield *et al.*, 1998; Xu, Grindley and Joyce, 2000). Similarly to eukaryotic FEN-1, Pol I 5'-nuclease threads the free 5'-end through this arch (Balakrishnan and Bambara, 2016). As riboflavin is slightly hydrophobic, there might be an interaction with Pol I 5'-nuclease arch such as to resist contact with FAD-flap, or interact too tightly to FAD-flap. Both ways would hinder 3'-5'-progression of the enzyme along DNA. Pol I also confers repair of apurinic/apyrimidic sites on damaged DNA (deoxyribose-phosphates (dRp) which lost their base) by displacing the dRp and the next one or two deoxyribonucleotides to excise small oligo“nucleotides” with 5'-dRp from the damaged strand (Price and Lindahl, 1991; Price, 1992).

RNase H was shown to remove primer RNA, but is non-essential if Pol I is present (Konrad and Lehman, 1974; Fukushima *et al.*, 2007; Lewis, Jergic and Dixon, 2016). Here, we tested if presence of 5'-NCIS on primer inhibits RNase H cleavage of primer. We observed that primer can be cleaved by RNase H, but resulting oligos might be longer than those of canonical primer (figure 4.12). Longer cleavage products were susceptible to NudC treatment, showing that RNase H likely does not cleave between NCIS and first canonical 5'-nucleotide. Sensitivity to NudC also shows that oligos dissociate from the template. These oligos have to be recycled by oligoribonuclease (Bandyra *et al.*, 2013). We intend to test oligoribonuclease activity on NCIS-capped oligos in the future.

Another potential downstream process of replication which has not been covered in this study is the primase-to-polymerase switch (or primer hand-over) on capped primer. In bacteria, the clamp loader complex loads β clamp dimer onto the RNA-DNA hybrid in a directional fashion for Pol III to bind and synthesise the Okazaki fragment (Kelch *et al.*, 2012). In yeast, it was proposed that the clamp loader recognises the primer-template junction via altered flexibility of the template in hybrid state vs ssDNA state (Bowman, O'Donnell and Kuriyan, 2004; Kelch *et al.*, 2012). Further, clamp loader might recognise typical RNA-DNA hybrid helix symmetry which is close to A-form DNA, while B-form DNA was shown to be excluded by clamp loader inner spiral chamber (Fedoroff, Salazar and Reid, 1993; Simonetta *et al.*, 2009; Kelch *et al.*, 2011). No study to date gives reason to assume the primer 5'-end matters for clamp loader binding. Although clamp loading likely recognises 3'-end of the primer, it might be worthwhile to test the effect of 5'-NCIN on clamp loading.

Bacterial primase DnaG is essential in the initiation of replication and its' inhibition is bactericidal (Ilic *et al.*, 2018). Study of primase inhibitors is thus of interest in development of new antibacterials. Interestingly, a list of DnaG inhibitors has been published (Ilic *et al.*,

2018) which are classed as NTP-analogues and non-NTP analogues. Here we see, that nucleotide-containing molecules with free 3'-OH can be utilised by DnaG to initiate primer synthesis efficiently. Greater understanding of DnaG structure and function provided by this study help screening for inhibitors, but use of primase inhibitors might also help us to understand NCIS-primer synthesis better. We intend to test several known DnaG inhibitors in the future.

4.5.3 Starvation response results in inhibition of NCIS-primer synthesis

Alarmones ppGpp and pppGpp in bacteria are produced by proteins RelA and SpoT in response to amino acid, fatty acid, carbon and iron starvation (Traxler *et al.*, 2008; Winther, Roghanian and Gerdes, 2018). While *B. subtilis* replication was shown to be directly inhibited by (p)ppGpp (Maciąg-Dorszyńska, Szalewska-Pałasz and Węgrzyn, 2013), it is unclear whether *E. coli* replication is inhibited (Maciąg *et al.*, 2010; Potrykus *et al.*, 2011). A direct interaction between *E. coli* DnaG and (p)ppGpp was shown *in vitro* (Maciąg *et al.*, 2010; Rymer *et al.*, 2012). Crystallisation studies of DnaG soaked in (p)ppGpp showed that the alarmone binding site is almost identical to nucleotide (GTP) binding site (Rymer *et al.*, 2012). As in (p)ppGpp, the 3'-OH is occupied by phosphate groups, chain extension is not possible. Thus competitive binding was proposed as mechanism of inhibition (Rymer *et al.*, 2012). Here we show, that ppGpp competes with cofactors for DnaG catalytic site, and inhibits primer synthesis in a concentration-dependent manner (figure 4.10). Due to the strong affinity of DnaG to ATP ($K_M=46\ \mu\text{M}$), primer synthesis is reduced by increasing concentration of ppGpp. Average stationary phase concentration of ppGpp around 150 μM (Varik *et al.*, 2017) reduced ATP-primer synthesis by less than 20%, but NADH-primer synthesis by over 40% and FAD-primer by over 70%. Maximal ppGpp of 0.9 mM (Traxler *et al.*, 2008) reduced ATP-primer synthesis by less than 60%, but both NADH- and FAD-primer over 90%. Thus, stationary growth *E. coli* do not seem to completely abolish primer synthesis, but incorporation of metabolites is inhibited. Higher ppGpp levels of stringent response inhibit NCIS-primer formation. This could be viewed as a mechanism to “save” valuable metabolites during starvation.

4.5.4 Eukaryotic primase might initiate primer synthesis with non-canonical substrates

NCIN-capping is performed by both multi- and single subunit RNAPs (bacterial RNAP, eukaryotic RNAP II, and mitochondrial POLMRT). We showed that also bacterial primase, which exhibits a different mechanism of RNA polymerisation, can initiate primer synthesis with abundant cellular metabolites. We thus wanted to investigate, whether eukaryotic

primase can similarly synthesise non-canonically initiated primer. Both enzymes are structurally and also mechanistically dissimilar, as described in the introduction.

Here we describe an initial test of this hypothesis which focussed only on the catalytic subunit of the two subunit primase, P49. P49 was previously shown to be active without P58, albeit at very low efficiency (Schneider *et al.*, 1998). P49 was purified from recombinant expression in *E.coli* in accordance to the protocol outlined by Schneider *et al.*, who describe stability and activity of the purified subunit is highly dependent on availability of divalent cation (Mg^{2+}) at every step of the purification. However, we have no positive control to compare the activity of our purified P49 to.

Our results of *in vitro* primer synthesis showed that P49 accepts m^7GTP and GpU as substrates to polymerise to $\alpha(P^{32})$ -UTP on an ssDNA template (figure 4.13). It is interesting that P49 accepted only a modified nucleotide and a dinucleotide as substrates, but not the canonical purine nucleotides. Theoretically, initiation requires assistance by large subunit P58 which contains the nucleotide binding site. Binding of the first two NTPs as well as DNA template are also required to fold the heterodimer into active conformation (Sauguet *et al.*, 2010). Apparently, altered binding characteristics of these two non-canonical substrates was able to stabilise the P49 active site enough to allow for catalysis. These results are encouraging to study cofactor incorporation in presence of P58 in the future. In the following text I would like to discuss theoretical considerations about NCIS initiation by P49P58, and theoretical implications on eukaryotic replication.

In archaea, PriX/L does not discriminate identity of initiating nucleobase or pentose and PriX affinity to ATP was found to be similar as to dATP. Thus, it was proposed that initiation of primer synthesis in archaea is regulated by substrate availability rather than template-dependent selection of substrate. NTPs are about 100 times more abundant in the cell than dTNPs, in favour of an RNA primer (Liew *et al.*, 2016; Yan *et al.*, 2018). In fact, the nucleotide-binding residues of the large primase subunit were to date described to interact only with 5'-triphosphate groups and not ribose or base (Sauguet *et al.*, 2010; Baranovskiy *et al.*, 2015). Without experimental evidence, we cannot assume that these conserved residues would bind NCIS, or constitute a regulatory mechanism against NCIS incorporation. The here observed ability of P49 to initiate primer at low efficiency with NCIS might indicate a previously unrecognised role of P49 in substrate binding. Interestingly, we have shown that P49 is inhibited by canonical substrates such as ATP and GTP, in contrast to results of (Schneider *et al.*, 1998). Possibly, the enzyme might bind canonical initiators, but lacks ability

to initiate polymerisation without P58. Studies on herpes primase showed that incorporation of dNTPs and NTP-ribose significantly shortened the final length of primer (Keller, Cavanaugh and Kuchta, 2008). Nucleotide analogues, such as 9- β -arabinofuranosyladenosine 5'-triphosphate (ara-ATP) were proposed to be potential eukaryotic primase inhibitors (Kuchta and Willhelm, 1991; Boudet *et al.*, 2015).

Eukaryotic and archaeal primase both contain FeS-cluster in the large subunit. Eukaryotic, but not archaeal, primase contains some sequence similarity to base-excision repair DNA Polymerase β . FeS-cluster are present in most eukaryotic replicative polymerases and are sensitive to nucleic acid electrochemical potential. The FeS cluster can be oxidised and reduced when bound to DNA and (d)NTPs (Beinert, Holm and Münck, 1997; Klinge *et al.*, 2007; Weiner *et al.*, 2007; Rouault, 2015; O'Brien *et al.*, 2017). P58 C-terminal domain (P58C) switches from resting redox state $[4Fe4S]^{+2}$ (loosely bound to DNA) to oxidised $[4Fe4S]^{+3}$, leading to 500-fold stronger affinity to DNA. This redox switching is considered a regulatory event for primer synthesis (Gorodetsky, Boal and Barton, 2006; O'Brien *et al.*, 2017; Tse, Zwang and Barton, 2017; O'Brien, Salay, *et al.*, 2018). It is possible, that presence of redox cofactors such as NAD(H) or FAD could affect this redox regulation. On the other hand, redox switching of primase P58C may regulate a distinction between NCIS of different redox states (e.g. NADH vs NAD⁺).

Further, Sheaff and Kuchta hypothesised that binding of penultimate NTP takes place before binding of 5'-NTP (Sheaff and Kuchta, 1993). This might explain why initiation with dinucleotide GpU was more successful than other G-analogues. Sauguet *et al.* compared the DNA and NTP binding site of yeast P58 to photolyase/cryptochrome flavoproteins, due to structural homology. The DNA photolyase cryptochrome DASH (a DNA repair enzyme) binds FAD as cofactor and photon recipient in repair of UV-mediated crosslinking mutations (Sauguet *et al.*, 2010). Both DNA and FAD binding sites of cryptochrome DASH and P58C have high similarity. Importantly, Sauguet *et al.* hypothesised the FAD binding site in DASH corresponds to dinucleotide binding site in P58 (Sauguet *et al.*, 2010). The involved FAD/NTP binding residues are conserved. It is thus conceivable that P58 may allow for binding of dinucleotide cofactors such as FAD or NADH during primer synthesis. We intend to test this hypothesis in the future.

Primer removal of prokaryotes and eukaryotes is not so different. Both involve a 5'-flap endonuclease, Pol I in bacteria, or FEN-1 in eukaryotes (Balakrishnan and Bambara, 2013). The arch structure in both enzymes is homologous (Robins *et al.*, 1994), therefore the

inhibitory effect of 5'-FAD on Pol I-mediated primer removal might turn out to be analogous in FEN-1.

4.6 Conclusion and outlook

In the presented study we report that bacterial primase DnaG efficiently initiates primer synthesis with adenosine-containing metabolites on recognition site 5'-d(CTG) *in vitro*. DnaG incorporates NADH and FAD efficiently, while DP-CoA and NAD⁺ were poor substrates. No sequential determinants of ssDNA template could be identified, but we showed that basic ridge residues of DnaG may be involved in NCIS incorporation.

We found that breakdown of primer by Pol I might be differentially affected. Interestingly, NADH-primer breakdown is more efficient than ATP-primer, while presence of 5'-FAD delays Pol I 5'-nuclease activity.

Initial experiments on human primase catalytic subunit P49 show that eukaryotes might be able to form NCIS-primer. Here, it is crucial to perform more experiments including the second primase subunit P58.

However and most paramount, verification of cofactor-primer *in vivo* still stands out. We are currently working on different approaches to investigate primer 5'-ends enzymatically, chemically and via liquid chromatography coupled with mass spectrometry (LC/MS). Various LC/MS methods are known for quantification of NCIN-caps on cellular RNA. Our prevalent difficulty is the enrichment of primer, which is a very short lived structure. We are attempting to enrich primer by growing double mutants of *rnhA* and *polA*.

5. Chapter III: Screen for Bacteriophage Regulators of *Mycobacterium smegmatis* Transcription

5.1 Introduction

Owing to the fast spread of antibiotic resistance and the increasing complexity of antibiotic treatment of mycobacterial infection (including TB), new treatment options need to be explored. Phage therapy is the administration of live phages to the infected patient in order to kill the bacterial pathogen. It has been known as alternative to antibiotics for almost a century and was used to treat many infections from cholera (D'Herelle, 1929) to chronic antibiotic-resistant otitis (Wright *et al.*, 2009). Application was also found in food preservation, agriculture and leather/pelt-production (Coffey *et al.*, 2011; Clark, 2015). Enzybiotics are a new group of protein enzymes of phage origin, which can lyse bacterial cells and have prospect of medical antibacterial application, as a bactericidal mechanism or to enhance susceptibility to other drugs (Nelson, Loomis and Fischetti, 2001; Dams and Briers, 2019). In response to intrinsic difficulties of anti-TB treatment, mycobacteriophages should be studied for their antimicrobial potential, both in phage therapy, and the newer field of enzybiotics. We think that also other bactericidal or bacteriostatic phage products, such as inhibitors of bacterial gene expression, could be used as therapeutic phage products. Further, investigating phage-host interactions provides us with a new pool of previously undiscovered inhibition mechanisms which are of interest in basic research of cellular processes. All information found in this introduction is also reviewed in (Puiu and Julius, 2019).

5.1.1 *Mycobacteriophages are diverse and resourceful*

Genomic studies have revealed that mycobacteriophages are among the most diverse class of bacteriophages so far characterised. Bacteriophages have evolved alongside their bacterial hosts, which consequently developed numerous defence mechanisms (Jacobs-Sera *et al.*, 2012). By developing spontaneous mutations, site-specific recombination and molecular mimicry (a form of autoimmune attack) phages can counteract bacterial defense (Cusick, Libbey and Fujinami, 2012). These processes have greatly influenced the genetic diversity of bacteriophages. Mycobacteriophage genomes present highly mosaic structures due to horizontal gene transfer and a wide evolutionary background (Hatfull, 2018). Approximately 1795 mycobacteriophage genomes have been sequenced to date, and arranged into clusters according to shared nucleotide sequence similarities: A total of 29 clusters, which share little or no sequence similarity, and five singletons (phages with no close relatives). Some of the

clusters are so diverse that they have been grouped into subclusters. The genomes in a cluster share at least 50% sequence similarity and features such as regulatory systems or tRNA sequences (Hatfull, 2014b, 2018). The potential to find new regulatory mechanisms in the mycobacterium-phage interaction is high.

Mycobacteriophages, as all bacteriophages, undergo two major types of life cycle: lytic and lysogenic (Young, Wang and Roof, 2000; Hatfull, 2014b) (see figure 5.1). Infection starts with adsorption of the phage to host cell receptor and injection of phage DNA into the cytoplasm. Lytic phages proceed to transcribe viral genes and replicate their genome. Subsequently, structural gene products (Gps) are organised into capsid (or ‘head’) and tail by assembly proteins. This is followed by association of viral daughter genomes and capsids, which leads to formation of virions. At the end of the lytic cycle, a phage-encoded holin forms pores in the bacterial cytoplasmic membrane, allowing phage endolysins to cleave the host cell wall, followed by the release of newly formed viral particles (Young, Wang and Roof, 2000). Temperate phages may opt for a lysogenic lifecycle, in which phage DNA integrates into the host chromosome where it remains as dormant prophage. In this case, phage integrase must insert the phage genome into the chromosome via homologous recombination (Jones and White, 1968). “Attachment” sites are present on the phage (*attP*) and bacterial genome (*attB*), usually located in tRNA genes (Hatfull, 2014a). After integration they form *attL* and *attR* sites, flanking the prophage (Peña *et al.*, 1997; Lewis and Hatfull, 2001). The majority of transcription of phage genes must be suppressed by a phage repressor protein and phage DNA is replicated along the host chromosome. A lytic cycle can be triggered by different stress conditions causing DNA damage. In some cases, phages may undergo a pseudotemperate lifestyle in which phage DNA does not integrate into the host chromosome, but replicates in the form of a plasmid in the host cytoplasm (Pope *et al.*, 2011).

To date, no mycobacteriophages were found to encode an own RNAP (Hatfull, 2018). Thus, mycobacteriophage transcription must exploit host RNAP. In many mycobacteriophages promoters corresponding to mycobacterial main sigma factor σ^A can be identified (Nesbit *et al.*, 1995; Hatfull, 2014a; Oldfield and Hatfull, 2014). However, other mycobacteriophage promoters are structured to be recognised by alternative host σ , or phage-encoded σ (Hatfull, 2014a). These phages regulate their transcription using their own gene products. Genome length is variable (40-160 kb) and usually divided into regions of structural (typically 20-25 kb, containing 20-40 genes) and non-structural genes (larger part of the genome). The non-structural genes encompass many open reading frames that are not yet identified. In fact,

many novel genes have been found in mycobacteriophage genomes that are unrelated to any GeneBank entries (Hatfull, 2014a).

5.1.2 Tuberculosis treatment requires research and innovation

Tuberculosis (TB) is an infectious disease of the lung or more rarely, of extrapulmonary tissues. In humans it is caused by *Mycobacterium tuberculosis* (Mtb) (Smith, 2003). Active tuberculosis lesions are caused by the immune response and bacterial growth. Depending on the tissue, these lesions can lead to further complications, such as hypoxia, if the lung is affected (Smith, 2003). The active stage of the disease is typically preceded by a latent phase of varying time span. Mtb can remain dormant in the host for weeks, months, years or even decades. After phagocytosis into macrophages or tissue cells, the bacteria are able to resist the phagolysosome, and the immune system proceeds to contain the infection by forming granuloma around infection foci. The granuloma are densely layered rigid structures meant to shield the body from the pathogen. Mtb can sense if the immune system is in a weakened state and liquefy the granuloma via an unknown mechanism (Silva Miranda *et al.*, 2012). The disease then progresses into the active stage. New cases of active pulmonary TB were documented in about 10 million people in 2017. 1.7 million died in 2016 (World Health Organization, 2018). Most of these cases are located within 30 high prevalence countries in the undeveloped world, due to lower standards of hygiene, poor economy and poor health care. Much lower prevalence in developed countries shows that high standard of living, including good nutrition and access to medical care, are the most important factors in the prevention of disease. However, outbreaks of TB, including drug resistant (DR) TB, are documented increasingly more often on the Northern hemisphere (Dye and Espinal, 2001; Udwadia and Vendoti, 2013; World Health Organization, 2018). People with autoimmune diseases, such as AIDS patients, are at higher risk of infection and development of acute disease, and fatality is higher in these subpopulations (Schutz *et al.*, 2010).

The treatment of TB as well as other diseases caused by mycobacteria (e.g. *M. ulcerans* ulcers, *M. leprae* leprosy) is particularly challenging due to the unique physiology and lifestyle of mycobacteria. Dormancy, intracellular infection and a thick cell wall with high lipid concentration render mycobacteria resistant to many treatments (Smith, 2003; Hoffmann *et al.*, 2008). Mycobacteria are also able to form biofilms, although it has not been determined whether this plays a role in TB (Kulka, Hatfull and Ojha, 2012; Brennan, 2017).

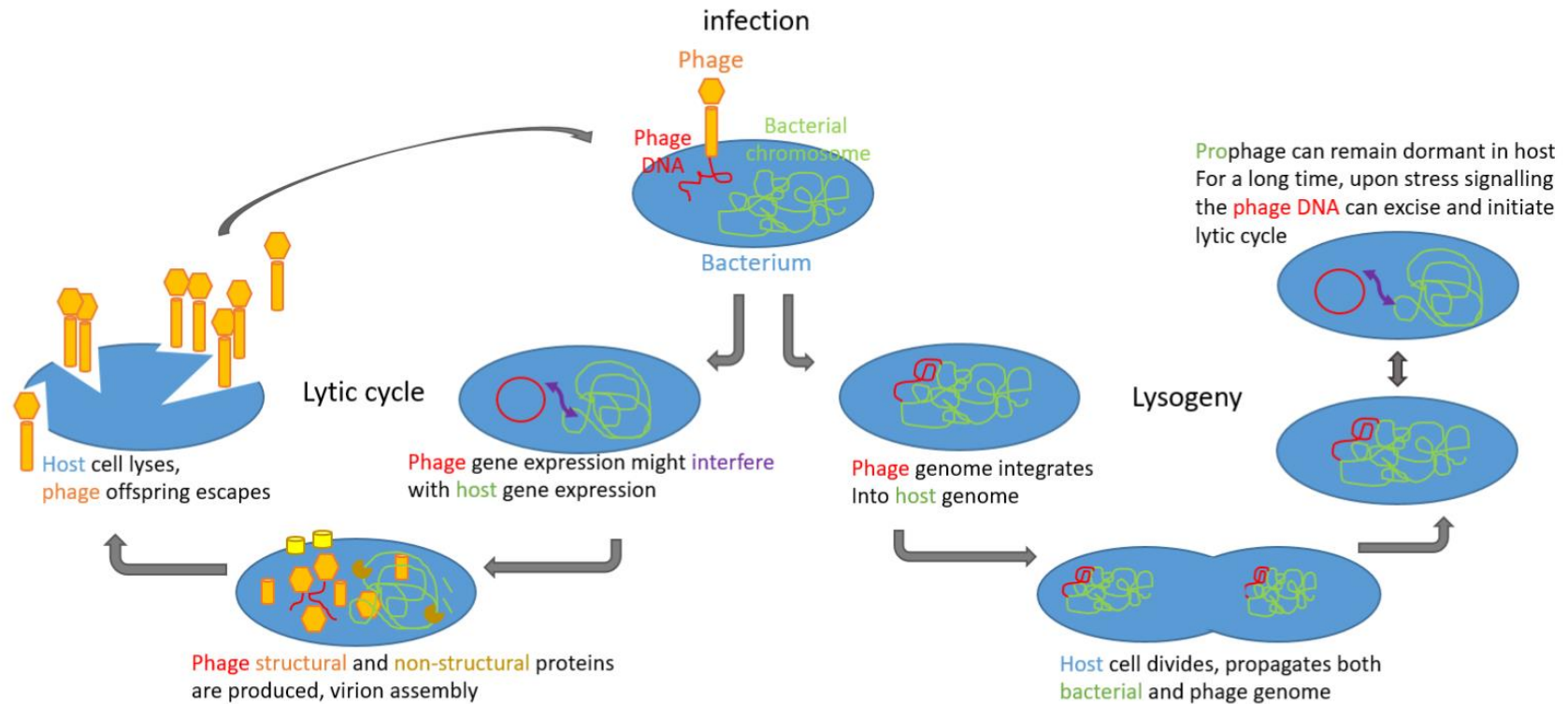


Figure 5.1 Phage lifestyles. Most phages are either lytic or lysogenic (temperate). This means that after infection, they either initiate phage gene expression and replication, followed by virion assembly, and finally lysis (lytic cycle). Or they integrate into the host genome, where they are maintained by host replication machinery until environmental signals lead to initiation of the lytic cycle (lysogeny).

Current treatment of drug susceptible (DS) TB constitutes a 6-month course of four drugs (rifampicin, isoniazid, ethambutol and pyrazinamide). The patient suffers side effects and damage to the natural microbiota (Smith, Wolff and Nguyen, 2013; Hofman *et al.*, 2016; World Health Organization, 2018). Although using a combination of drugs and long treatment duration reduces the chances of resistance development, DR-TB is on the rise (Zumla *et al.*, 1999; Tiberi *et al.*, 2018). TB is considered drug-resistant if sensitivity to at least one drug of the regimen is lost. Most commonly, rifampicin resistance develops due to chromosomal mutations in rif-binding pocket of RNAP (Koch, Mizrahi and Warner, 2014). If the strain is resistant to at least two drugs of the first line treatment, it is considered multidrug resistant (MDR). Roughly 4.5 % of new cases of active TB reported in 2015 were MDR (Pai *et al.*, 2016). Most commonly resistance to rif and isoniazid are reported (Hofman *et al.*, 2016). In such cases, second-line and third-line treatments are applied which can include further drugs which are currently in clinical trials (e.g. bedaquiline, delamanid). MDR treatment has higher toxicity, cost and longer duration. Resistance to third-line or all drug treatments have been reported. These strains are referred to as extensively drug resistant (XDR) and pan resistant, respectively (Stoffels *et al.*, 2013; Udwadia and Vendoti, 2013; Hofman *et al.*, 2016).

Survival of active DS-TB is estimated to 86% with first line-treatment. MDR-TB treatment is successful in 50% of cases (Blumberg *et al.*, 2003; Winston and Mitruka, 2012). The WHO is motivated to support research on alternative anti-TB strategies (World Health Organization, 2018).

Currently, treatment alternatives are mainly used in cases where standardised treatment is not available (poor and rural areas), and often includes traditionally used medicinal herbs. Some of these herbal treatments have been documented as effective (Semenya and Maroyi, 2013; Nguta *et al.*, 2015; Sharifi-Rad *et al.*, 2017), and could be investigated in search of plant-derived drugs. Immunotherapy can be used to support antibiotic treatment. Administration of cytokines and other antigens (such as hepatitis virus antigens) or interferons was shown to modulate the immune response to reduce tissue damage and promote pathogen clearance (Batdelger *et al.*, 2008; Dawson *et al.*, 2009; Arjanova *et al.*, 2011; Butov *et al.*, 2013; Hofman *et al.*, 2016). TB vaccination is currently administered by regulation in the largest part of the world (Asia, Eurasia, Africa, South America) (Zwerling *et al.*, 2011), but in North America and Western Europe it is mainly administered to health care workers and people with increased risk, e.g. certain indigenous Canadians or people in MDR endemic regions (Vaudry, 2003). This vaccine contains an attenuated strain of *M. bovis* called Bacille Calmette-Guerin

(BCG). Unfortunately, it shows varying degrees of effectiveness, between 0-83%, in treatment of the disease, depending on age and genetic background of the patient, geographic location and climate, Mtb strain variations and other uncontrollable factors (Vaudry, 2003). A variety of vaccine boosters and new preventive vaccines are under study (Hofman *et al.*, 2016).

5.1.3 *Mycobacteriophage therapy could aid TB treatment*

All potential new drugs must overcome the challenges of tuberculosis infection: Anti-TB drugs must be able to cross the mycobacterial membrane, the mammalian cell membrane, penetrate several types of tissue, most importantly granuloma, and should not interfere with HIV therapy. Furthermore, short treatment duration and low toxicity are just as desirable as low adverse effect on the microflora and good bioavailability. Additionally, potential to be distributed to remote and poor areas would be beneficial. Phage therapy has advantages and disadvantages. Bacteriophages have low production cost and virtually no effect on bacteria which are not their natural host. No direct adverse interaction of phages with drugs have been documented. Phages can invade their host naturally by interacting with specific receptors on the cell envelope (Bertozzi Silva, Storms and Sauvageau, 2016), and dosage adjusts naturally with availability of live host cells, as phages multiply inside the host cell (Bessman *et al.*, 2001; Capparelli *et al.*, 2007; Chhibber, Kaur and Kumari, 2008). However, phages do not cross mammalian cell membranes, let alone multiple tissue layers (Mankiewicz and Béland, 1972; Bessman *et al.*, 2001). They can be removed by the animal immune system and can create immune memory (Nieth *et al.*, 2015; Singla *et al.*, 2016; Krut and Bekeredjian-Ding, 2018) which could make long-term or repeated treatment obsolete. They are usually stored in liquid suspension and may require refrigeration, which could be an obstacle for distribution in poor countries with hot climate. The problem of clearance of phage by the immune system may be overcome by using liposomes, as shown for *Klebsiella pneumoniae* intracellular infection (Singla *et al.*, 2016). Further, some scientists argue that while phage will be attacked by the immune system, their ability to multiply within the bacterial host will maintain phage titre while aiding bacterial clearance by modulating the immune response. This is called “immunophage synergy” (Leung and Weitz, 2017; Roach *et al.*, 2017). Likewise, phage-antibiotic synergy (PAS) has been documented. Here, phage can potentiate drug delivery by making the bacterial cell wall and even biofilm more permeable. Pre-existing drug resistance can even be reversed by PAS (Ryan *et al.*, 2012; Jo, Ding and Ahn, 2016; Kim *et al.*, 2018; Tkhilaishvili *et al.*, 2018).

Mycobacteriophages in particular have been applied successfully for TB treatment *in vitro* and in animal models. Bacterial titre was reduced in the animal model of Mtb lung, liver and spleen infection (Sula, Sulová and Stolcpartová, 1981) and *M. ulcerans* ulcer on feet (Trigo *et al.*, 2013). Even granuloma formation was reduced (Sula, Sulová and Stolcpartová, 1981). Still, transport across the animal cell membrane is an obstacle. Broxmeyer *et al.* developed a delivery system of phage to Mtb across the macrophage membrane using *M. smegmatis* as a vehicle. Mycobacteriophage TM4 can infect both Mtb and *M. smegmatis*. TM4-infected *M. smegmatis* were phagocytosed by macrophages which already carried Mtb infection. This resulted in infection of Mtb by TM4 and reduction of titre of both bacteria within the macrophage (Broxmeyer *et al.*, 2002).

5.1.4 Enzybiotics are protein antibacterials of phage origin

Phage therapy encounters many challenges which research needs to overcome. Additionally, acceptance of phage therapy in Western medicine is historically low. Although phage therapy has been in use since before first antibiotics were discovered (Sulakvelidze, Alavidze and Morris, 2001), its effectivity is not widely accepted. Further limitations (high pathogen-specificity, limited tissue permeation, standard hygienic requirements of drug manufacturing) did not make phage therapy a promising research topic to many scientists and funders. However, phage Gps have high potential to provide us with new bactericidal and bacteriostatic mechanisms, which could potentially be implemented in new TB treatment regimen.

Enzybiotics is a new field that is being researched since the early 2000s and focusses on lytic enzymes encoded by bacteriophages (Nelson, Loomis and Fischetti, 2001). At the end of the lytic lifecycle of infection, the host cell wall is lysed via a two-step process: Phage endolysin is expressed at low level and accumulates during phage growth. Phage holin is expressed only in late infection, and creates inner membrane channels through which endolysin escapes into the periplasmic space. It then lyses peptidoglycan, which allows for virus progeny to rupture the cell (Schmelcher, Donovan and Loessner, 2012). Phage endolysins, although the name correctly suggests they lyse cells from the inside, were successfully applied exogenously to lyse cells from the outside (Nelson, Loomis and Fischetti, 2001; Schmelcher, Donovan and Loessner, 2012). Phage endolysins were shown to be highly specific to their bacterial host (Nelson, Loomis and Fischetti, 2001; Bustamante *et al.*, 2010). Pre-existing antibiotic resistance likely does not affect phage-derived drugs. What is more, PAS could potentially be exploited using enzybiotics, as they can make biofilm and the bacterial cell envelope more

permeable to drugs and thus potentiate antibiotic treatment (Hermoso, García and García, 2007; Manoharadas, Witte and Bläsi, 2009; Domingo-Calap and Delgado-Martínez, 2018; São-José, 2018). Enzybiotics can be engineered to facilitate delivery of the drug to the (intracellular) bacterial pathogen. In an effort to overcome the animal cell barrier, Wang *et al.* engineered a fusion protein of phage endolysin to a murine transcription activator. This transcription factor can cross the murine cell membrane, fused to endolysin, to treat intracellular *S. aureus* infection (Wang *et al.*, 2018). Also recombinant expression of phage endolysin was shown to be effective in prevention of *S. aureus* infection in transgenic cattle and mice (Cheng *et al.*, 2005; Wall *et al.*, 2005).

All to date investigated mycobacteriophages encode at least the two lysins, peptidoglycan endolysin LysA, and lipolysin LysB (Gil *et al.*, 2008; Mahapatra *et al.*, 2013; Lai *et al.*, 2015). Mycobacteriophages Ms6 and BTCU-1 were shown to produce lytic enzymes which can lyse *M. smegmatis* bacterial culture as well as intracellularly (Gil *et al.*, 2008; Catalão *et al.*, 2011; Mahapatra *et al.*, 2013; Lai *et al.*, 2015; Catalão and Pimentel, 2018). A large screen of 200 synthetic proteins generated from a mycobacteriophage database helped to identify several anti-mycobacterial peptides. Among these was PK34, a protein that binds Mtb glycolipid, and was successfully applied to clear TB in the mouse model (Wei *et al.*, 2013; Teng, Liu and Wei, 2015). It is worthwhile to look for further antibacterials from mycobacteriophages.

We propose that, besides endolysins, other mycobacteriophage proteins might be useful in phage anti-TB treatment, and could fall under the umbrella of enzybiotics.

Mycobacteriophages are diverse and relatively well characterised. Several studies investigating mycobacteriophage L5 have yielded a variety of potentially cytotoxic proteins, some of which were shown to downregulate *M. smegmatis* gene expression (Donnelly-Wu, Jacobs and Hatfull, 1993; Hatfull and Sarkis, 1993; Rybniker *et al.*, 2008). Few antibacterial mycobacteriophage effectors of bacterial transcription have to date been investigated. One is TM4 encoded transcriptional regulator of mycobacterial cell division, TM4-WhiB (Rybniker *et al.*, 2008). We aim to identify further inhibitors of mycobacterial gene expression with the method outlined in this chapter.

5.1.5 *Host transcription regulation is a common phage strategy*

Inhibition of host gene expression during viral infection is called host shut-off (McAllister and Barrett, 1977; Roucourt and Lavigne, 2009). If a phage does not encode its own RNAP, it might not completely shut off cellular transcription, but rather manipulate it. Many phages are able to downregulate host transcription to the benefit of their own gene expression, to monopolise resources (NTPs, enzymes, tRNAs, etc) and prepare host lysis (e.g. abolish repair mechanisms). Additionally, they employ mechanisms to redirect host RNAP to prioritise viral promoters.

Thus far, no mycobacteriophage has been described which encodes an own RNAP (Hatfull, 2014a). For this reason, it can be assumed that mycobacteriophages manipulate host RNAP to express phage genes. RNAP gene sequences are vastly conserved throughout the prokaryotes and altogether about 25% of sequences show evolutionary variability. Variable regions are mostly expendable and their deletion does not affect basic enzymatic activity. However, they provide targets for species-specific regulators as well as exogenous interaction partners, like phage encoded regulators (Nechaev *et al.*, 2002; Nechaev and Severinov, 2003). As mycobacteria have been proposed to have gone through many host transitions in the past (Hatfull, 2008), RNAP conserved regions also make a good target requiring no adaption in a new host. Therefore, we think that transcriptional regulators can be found via exploring phage proteins that interact with mycobacterial RNAP.

The process of transcription in bacteria was described in the general introduction. In short, it can be segregated into 6 main steps, all of which could be potential targets of interference by phage Gps. i) Holoenzyme formation and promoter recognition, ii) Promoter melting and isomerization from closed to open promoter complex, iii) Transcription initiation, iv) Promoter escape, v) Elongation, and finally vi) Transcription termination (Ma, Yang and Lewis, 2016). The RNAP core is made up of the protein subunits α , β , β' , and ω , completed by the σ factor to give the holoenzyme. The promoter DNA is an essential determinant of transcription as binding of σ is sequence and structure specific (Kalate, Kulkarni and Nagaraja, 2002). In mycobacteria, transcription factor RbpA was determined to be necessary for transcription initiation on many promoters (Tabib-Salazar *et al.*, 2013; Perumal *et al.*, 2018). Mycobacterial promoters differ from *E. coli* promoters in many ways. While a conserved -10 element of the sequence TATAaT (small a is found in less than 50% of mycobacterial promoters) was identified in *M. smegmatis*, -35 region consensus remains undetermined (Newton-Foot and Gey Van Pittius, 2013). This is likely due to high G/C

content of mycobacteria, as well as the wide variety of σ factors (*M. smegmatis* has 26 σ) (Mulder *et al.*, 1997; Newton-Foot and Gey Van Pittius, 2013). Instead, additional sequence motifs constitute binding sites for alternative σ and other, environmentally regulated, transcription factors (Newton-Foot and Gey Van Pittius, 2013). Strongest *M. smegmatis* promoters have an extended -10 element (TGn). RbpA binds σ^A and σ^B promoters and substitutes for missing -35 and extended -10 TGn motifs (Perumal *et al.*, 2018). Mycobacterial RNAP can perform intrinsic termination in absence of RNA termination loops, meaning only via 7-9 nt U-tracts, in presence of elongation and termination factor NusG (Czyz *et al.*, 2014). Also, Rho-dependent termination is described in mycobacteria (D'Heygère *et al.*, 2015; Botella *et al.*, 2017).

Not much is known about mycobacteriophage transcription. Lysogenic mycobacteriophage L5 gene expression is susceptible to rifampicin treatment, which is an argument in favour of L5 utilising bacterial RNAP. L5 gene expression results in shut-off of host gene expression (Hatfull and Sarkis, 1993; Hatfull, 2014a). However, for mycobacteriophage TM4 no host shut-off was observed (Ford, Stenstrom, *et al.*, 1998). TM4 is pseudotemperate (Ford, Stenstrom, *et al.*, 1998), meaning its genome is maintained in the host cell in plasmid form. If TM4 uses host RNAP without inhibition of host gene expression, the cell can remain healthy and provide factors until lysis is initiated. TM4, like many phages, regulates its own gene expression in chronological phases. Early phase genes are expressed upon activation of lytic cycle or 20-60 minutes post infection (p.i.), late phase after 60 minutes p.i. (Ford, Stenstrom, *et al.*, 1998). Figure 5.2 shows transcription of a hypothetical mycobacteriophage that, like TM4, regulates its own gene expression in early and late phases. Late genes often code for structural and assembly proteins. Mycobacteriophage SWU1 ensures its own gene expression via an alternative route, by enhancing overall cellular gene expression. SWU1 regulates host transcription in a differential fashion. While some genes are downregulated (e.g. siderophores), genes of transcription (RNAP subunits), translation, RNA degradation and protein transport and secretion are upregulated (X. Fan *et al.*, 2016). Yet another phage trick to ensure its own transcription is the use of particularly strong promoters, as shown for BP (Oldfield and Hatfull, 2014). Transcription termination is targeted by mycobacteriophage Ms6 encoded anti-termination factor (Garcia, Pimentel and Moniz-Pereira, 2002). Anti-termination is employed by many phages to facilitate transcription of long polycistronic operons (Gottesman and Weisberg, 1995; Santangelo and Artsimovitch, 2011). The hypothetical mycobacteriophage of figure 5.2 uses anti-termination to transcribe long polycistronic RNAs.

Table 5.1 lists different mechanisms of transcription regulation by various phages as well as drugs. Phage regulators are proteinaceous while antibiotics are usually smaller organic compounds (with exception of Microcin J25, which is a peptide). It becomes apparent that phage regulators usually do not target the same epitope, or employ the same inhibition mechanism as drugs.

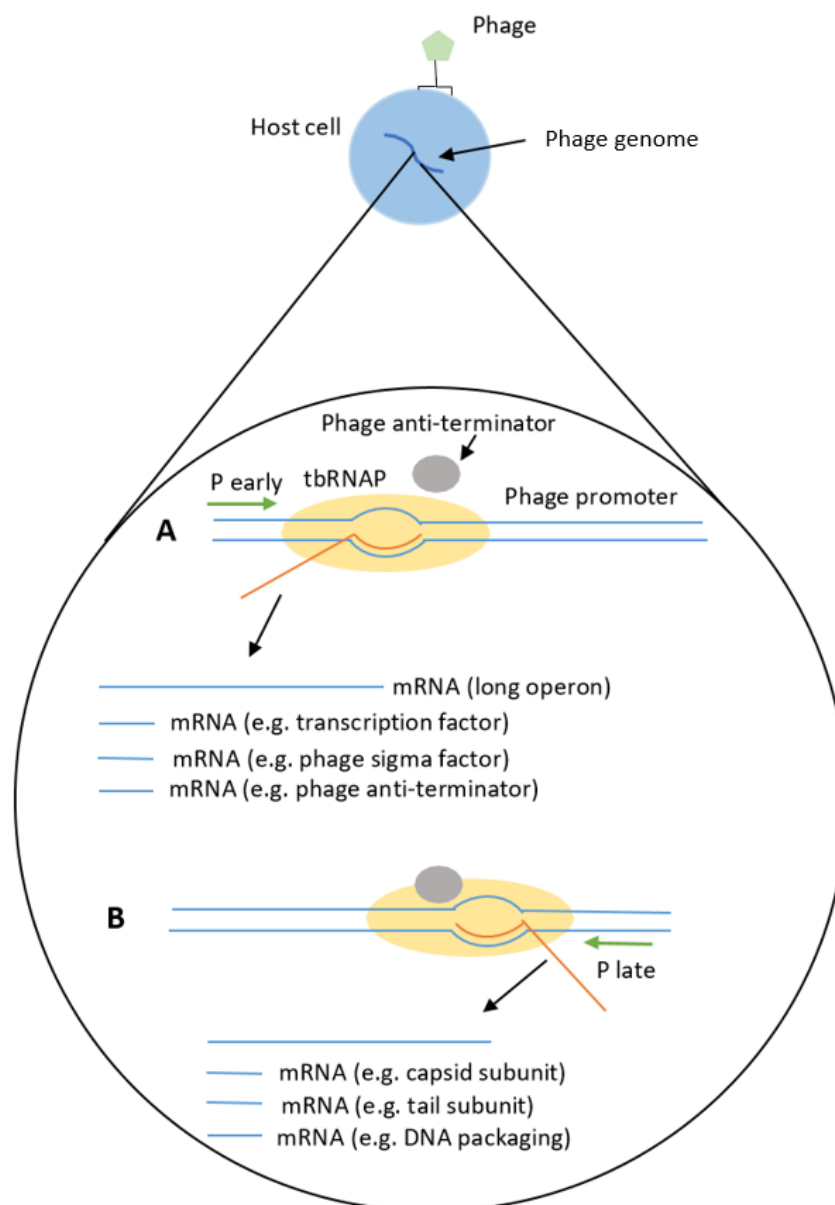


Figure 5.2 from (Puiu and Julius, 2019). Gene expression by a hypothetical lytic mycobacteriophage. The mycobacteriophage genome is transcribed by mycobacterial RNAP (tbRNAP). Transcription of this phage is regulated into early (A) and late (B) stage of infection, yielding long polycistronic transcripts. Early genes often encode non-structural proteins such as regulators of phage and host gene expression, e.g. phage antiterminator. Antiterminator is a common regulatory mechanism employed by phages to ensure transcription of promoter-distal genes. Late genes typically encode structural proteins or proteins of functions in virus particle assembly (Hatfull, 2014a).

Table 5.1 Inhibitors of bacterial transcription of phage origin (phage effector) vs antibiotics or chemicals that are used in clinic or research. For each stage of transcription, phage Gps and drugs are listed with name, function and reference.

Target	Phage effector		Other inhibitory substances	
	Protein (phage, host)	Function (reference)	Inhibitor (classification)	Function (reference)
Initiation				
σ binding	AsiA (T4, <i>E. coli</i>)	Anti- σ / σ appropriation (Orsini <i>et al.</i> , 1993; Pal <i>et al.</i> , 2003; Lambert <i>et al.</i> , 2004)	Compound C5 (developmental drug)	Blocks β' σ binding site (Ma, Yang and Lewis, 2016)
	G1ORF67/ Gp67 (G1, <i>S. aureus</i>)	Anti- σ (Dehbi <i>et al.</i> , 2009)		
	Gp39 (P23-45, <i>T. thermophilus</i>)	σ -appropriation (Tagami <i>et al.</i> , 2014; Ooi <i>et al.</i> , 2018)		
Promoter binding	AsiA + MotA (T4, <i>E. coli</i>)	Redirection of RNAP to phage promoter (Guild <i>et al.</i> , 1988; Ouhammouch <i>et al.</i> , 1995; Hinton <i>et al.</i> , 1996)		
	G1ORF67/ Gp67 (G1, <i>S. aureus</i>)	Redirection of RNAP from UP-element host promoters to phage promoters (Dehbi <i>et al.</i> , 2009; Osmundson and Darst, 2013)		
Open complex formation	Gp2 (T7, <i>E. coli</i>)	Destabilises RPo (Nechaev and Severinov, 1999; Klimuk <i>et al.</i> , 2013)	Fidamoxcin (antibiotic)	Destabilises RPo via σ and RbpA (Fruth <i>et al.</i> , 2014)
	Gp76 (P23-45, <i>T. thermophilus</i>)	Redirection of RNAP from host -35/-10 to phage extended -10 promoters (Berdygulova <i>et al.</i> , 2011; Ooi <i>et al.</i> , 2018)	Myxopironin (antibiotic candidate)	Blockage of switch region (Fruth <i>et al.</i> , 2014)

Initiation catalysis			Rifampicin (antibiotic)	Blockage of RNA polymerisation >3 nt (McClure and Cech, 1978; Campbell <i>et al.</i> , 2001b)
			Kanglemycin (antibiotic candidate)	Interferes with phosphodiesterbond formation (Mosaei <i>et al.</i> , 2018)
			GE23077 (antibiotic candidate)	Binds active centre, prevents NTP binding (Zhang <i>et al.</i> , 2014)
Elongation				
Ternary elongation complex catalysis			Salinamide (antibiotic candidate)	Allosteric alteration of active site (Degen <i>et al.</i> , 2014)
			Streptolygidin	Stabilises inactive intermediate of TEC (Temiakov <i>et al.</i> , 2005)
Pausing	P7 (Xp10, <i>X. oryzae</i>)	Stabilisation of TEC to inhibit pausing (Zenkin, Severinov and Yuzenkova, 2015)	Microcin J25 (peptide antibiotic)	Blocks secondary channel (Adelman <i>et al.</i> , 2004)
Termination				
Termination	N-protein (λ , <i>E. coli</i>)	Binding of elongation factors creates termination resistant TEC (Roberts, 1969; Yang <i>et al.</i> , 2014)		
	p7 (Xp10, <i>X. oryzae</i>)	Binding of β' stabilises TEC (Yuzenkova <i>et al.</i> , 2003)		
	Gp39 (P23-45, <i>T. thermophiles</i>)	Stabilisation of TEC to transcribe through poly (U) termination signal (Berdygulova <i>et al.</i> , 2012)		

5.2 Aims

We intend to identify new phage effectors of an essential cellular process, transcription. We attempt to screen for all potential phage Gps which regulate transcription via direct interaction with RNAP. Using a mutant strain with His-tagged chromosomal RpoC (β' -subunit) (provided by Prof Dipankar Chatterji, Indian Institute of Science, India) we aim to sequester phage proteins at different time points during infection. The mix of associated phage proteins can be identified by mass spectrometry, expressed recombinantly, and then investigated individually.

5.3 Results and Discussion

5.3.1 Workflow

This method was developed using three different mycobacteriophages, D29, TM4 and Adephagia $\Delta 41\Delta 43$ (lytic mutant of temperate phage Adephagia). Because we ended up finding a promising inhibitor of mycobacterial transcription from D29, figures presented here depict results from D29.

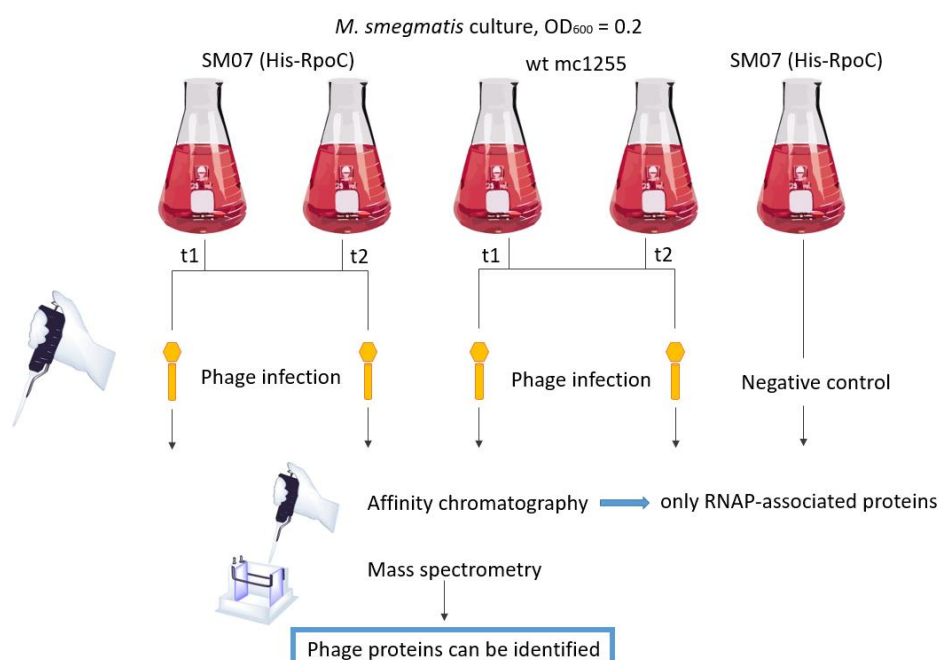


Figure 5.3. Schematic of infection workflow. Five liquid cultures of *M. smegmatis* SM07 or wildtype (wt) mc1255 are simultaneously infected with mycobacteriophage. Infected cells are harvested at two different timepoints post infection (t1 and t2). Cells are lysed and the lysate is subjected to affinity chromatography that is specific to the RNAP of strain SM07 (His-RpoC). All proteins that are purified due to their interaction with His-tagged *M. smegmatis* RNAP are identified by mass spectrometry. Only those phage proteins which are contained in SM07-infected samples, and not wildtype control or not-infected control, are cloned and recombinantly expressed for further analysis.

In order to screen for phage Gps that interact with RNAP, we prepared several samples and controls, as outlined in figure 5.3. *M. smegmatis* SM07 (His-tagged RpoC) or as method control wildtype *M. smegmatis* mc1255 liquid cultures are infected with mycobacteriophages at two different time point post infection, t1 and t2. As negative control, no infection occurs. RNAP from cell lysates is purified, and proteins that can be pulled down via RpoC are identified via mass spectrometry. Then, identified phage proteins were cloned, recombinantly, expressed, and investigated.

This protocol and additional procedures for optimisation are described in the following text.

5.3.2 Optimisation of *M. smegmatis* growth conditions

We intended to observe *M. smegmatis* growth in order to develop a protocol that can be applied in a standardised way. This method was outlined in reference to (Hatfull and Jacobs, 2014; Parish and Roberts, 2015) and private conversation with Deborah Jacobs-Sera from the Hatfull Lab, Pittsburgh University, USA.

First of all, uniform infection requires knowing the infectious dose, or multiplicity of infection (MOI). This number describes the ratio of phage to bacterium that is mixed. In mycobacteriophage protocols, traditionally, an MOI of 10-100 is used (10-100 phages per bacterial cell) (Hatfull, 2014a). In order to guarantee a known MOI we first determined the correlation of OD₆₀₀ and cell count by dilution series and colony counting. For this, we sampled *M. smegmatis* cultures at several ODs, performed a serial dilution of the sample and plated the dilutions on solid medium.

With knowledge of dilution and applied volumes we can calculate the concentration of colony forming units (cfu)/ml for each OD after counting colonies on agar plates. OD₆₀₀ = 0.1 corresponded to $2.5 \cdot 10^7$ cfu/ml. OD₆₀₀ = 0.2 corresponded to $5 \cdot 10^7$ cfu/mL.

To ensure a smooth process we determined growth times at different temperatures. We determined that at 30°C *M. smegmatis* culture inoculated 1:1000 (e.g. 100 µl starter culture in 100 ml medium) grows to OD₆₀₀ = 0.2 within 24 h. *M. smegmatis* grows to OD₆₀₀ = 1.0 within 24 h at 37°C (see figure 5.4). Thus, cultures to be infected were inoculated 24 h prior infection and grown at 30 °C to achieve the desired OD. We further observed that incubation in light induces yellow pigmentation which has been known to be caused by utilisation of alternative σ^F (Provvedi *et al.*, 2008). In order to standardise composition of RNAP holoenzyme, we incubated cultures in the dark (as far as this is possible in a communal incubator).

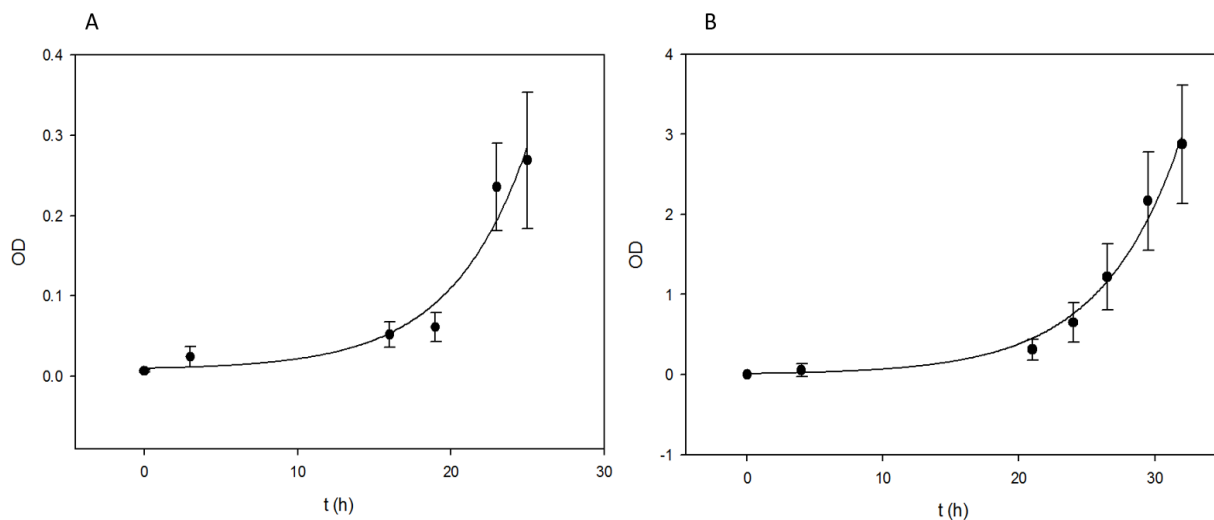


Figure 5.4 Growth of *M. smegmatis* in liquid culture at different temperatures. Cells were inoculated 1:1000 and grown in 7H9 broth supplemented with 10% ADC supplement, 0.02% Tyloxapol and 50 µg/ml carbenicillin at 30 °C (A) or 37 °C (B). At 30°C culture grows to OD₆₀₀=0.2 within 24 h, and at 37°C to almost 1.0.

A common problem encountered when growing *M. smegmatis* in liquid culture is clumping. Usually, clumping is prevented by addition of detergent, here tyloxapol (0.02-0.04%). Detergent cannot be applied during infection, as it interferes with phage adsorption. We thus tested D-xylose as alternative against cell-aggregation which was suggested to reduce clumping (Anton, Rougé and Daffé, 1996). We have previously observed that culture clumping translates to wide error between OD measurements. Thus Figure 5.5 shows measurement error between 4 cultures of *M. smegmatis* grown in 0.02 % tyloxapol or 1 mg/l D-Xylose over 24 h. We deduct that tyloxapol and xylose prevent clumping to a similar extend. Thus we added 1 mg/L xylose to cultures when tyloxapol had to be omitted.

5.3.3 Phage titration methods

These protocols were designed in reference to (Hatfull, 2010a) and information found on Actinobacteriophages Database website (Hatfull, Graham, Russell, Dan, Jacobs-Sera, Debbie, Pope, Welkin, Sivanathan, Viknesh, Tse, 2016).

For infection with known MOI also the phage titre needs to be known. For titration of phage, a phage suspension is serially diluted and added to a sample of cells. After 30 min adsorption, the infected cells are mixed into liquid top agar (0.5 %) agar and poured onto solid medium. Bacterial lawn grows overnight and infected cells lyse, but phage diffusion through top agar is limited, thus plaques (foci of lysed cells) can be counted to calculate the plaque forming units (pfu)/ml of original phage stock.

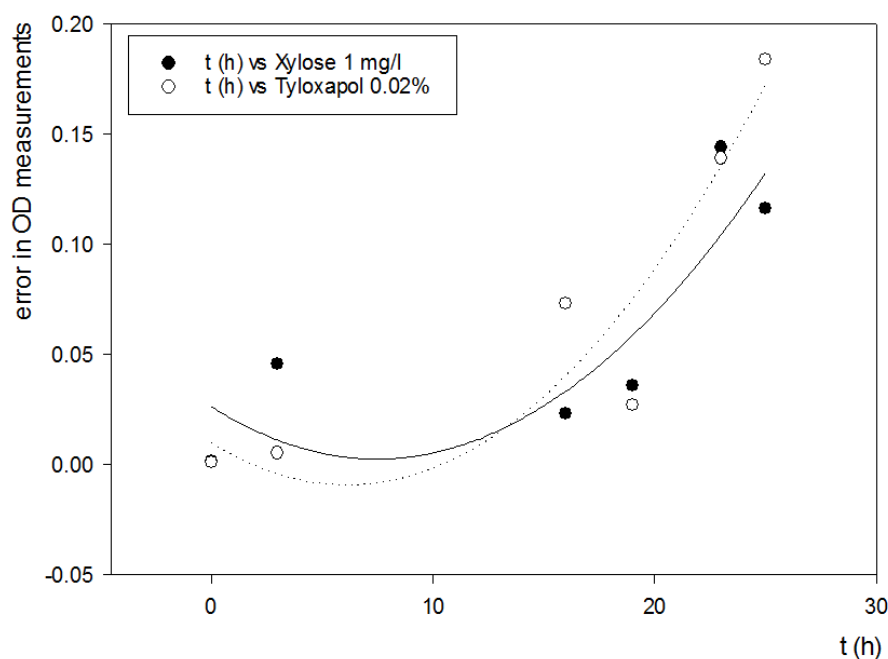


Figure 5.5 Evaluation of tyloxapol or xylose to prevent *M. smegmatis* cell aggregation. Traditionally, detergents are added to liquid cultures of *M. smegmatis* to prevent cell aggregation. This is not possible in cultures which are grown to be infected by bacteriophage as detergent inhibits phage adhesion to the cell. Different sugars were indicated to prevent clumping of mycobacteria (Anton, Rougé and Daffé, 1996). By plotting the error in OD₆₀₀ between measurement of 4 cultures grown in parallel in 1 mg/L D-xylose (black dots and full line quadratic standard curve) and 0.02% tyloxapol (white dots and dotted quadratic standard curve) we can see that degree of clumping between both methods is similar.

Spot test is similar to previously described titration, but here, the bacterial lawn is poured first, then phage dilutions are dropped onto the top agar. From the dilutions which show a countable number of plaques within the drop, the pfu/ml can be calculated.

Figure 5.6 shows spot test and whole plate assays for phage titration. Both methods require serial dilutions of phage to be plated on a bacterial lawn. If single plaques can be counted, pfu/ml can be calculated via the below calculation.

$$phage\ titre\ \left(\frac{pfu}{mL}\right) = \frac{\#plaques * dilution\ factor * Volume\ factor\ (e.g.\ 10^3\ for\ \mu L)}{Volume\ (e.g.\ 10\ for\ 10\ \mu L)}$$

5.3.4 Determination of an infection timeline

Cells are synchronised prior infection. Synchronisation is achieved by inoculating the infection culture with a pre-culture in early stationary phase (growth at 37 °C for at least 24 h) and by pelleting cells prior infection (Hoppensteadt, 1989; Kirtania *et al.*, 2016). The pelleted

cells are resuspended in phage buffer which provides ideal ionic strength for phages to adsorb to the cell. Adsorption is done in small volume in absence of carbon source and infection ($t = 0$) is initiated by adding fresh growth medium.

Alternative methods of synchronisation, such as use of vitamin C, have been described (Kirtania *et al.*, 2016), but not tested by us. In order to determine time points of infection in which we can assume phage gene expression is in progress, we observed the correlation between cell lysis (measured by OD) and phage titre (measured by spot test).

If measurements of OD and phage titre are plotted (figure 5.7) we can see that a steep fall of OD coincides with a steep rise in phage titre, as expected. In case of phage D29 this takes place about 110 min p.i. Therefore we chose to harvest infected cells 15 and 40 min p.i. for investigation of early and late Gps ($t = 0$ is after 30 min of adsorption). Infection dose was MOI 10 (ratio of cfu to pfu 1:10). MOI 100 was also tested (data not shown). As the resulting curve did not vary noticeably, we chose to proceed with MOI 10.

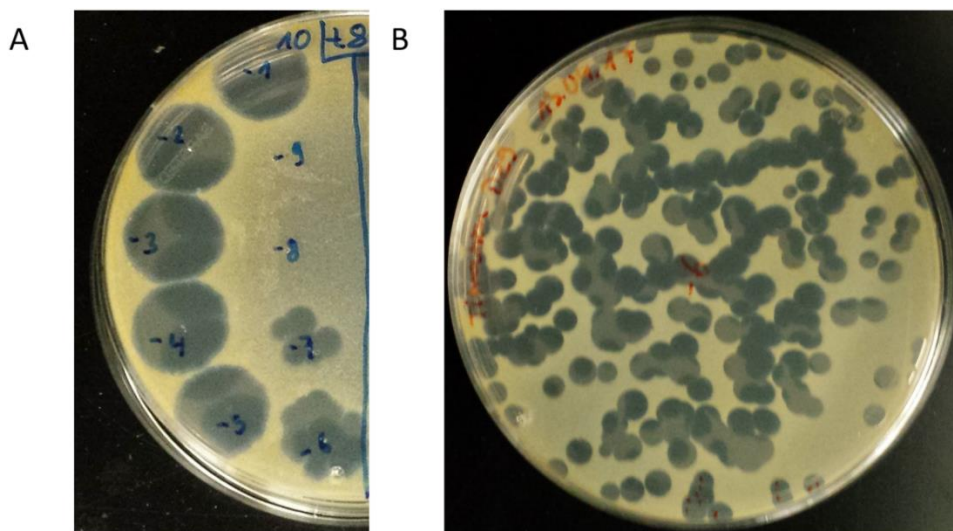


Figure 5.6 Spot test and plate assay for titration. Phage concentration in a suspension is calculated by counting plaque forming units after serial dilution. A. Spot test. Bacterial lawn is poured onto an agar plate, then 10 μ L drops of dilutions of phage suspension are dropped onto the freshly inoculated lawn. Bacteria grow in absence of phage, phage forms holes (plaques) in the lawn. Here, dilution 10^{-7} allows for counting of 4 plaques, with the above equation we determine a phage titre of 4×10^9 pfu/mL. B. Plate assay. Here, instead of dropping phage onto the lawn, the phage (10 μ L of one dilution, here 10^{-6}) is dispersed into cell suspension before the top agar is poured. The bacterial lawn is laced with plaques. Here, 210 plaques were counted, giving a titre of 2.1×10^{10} pfu/mL. Both plates show D29, but not the same samples.

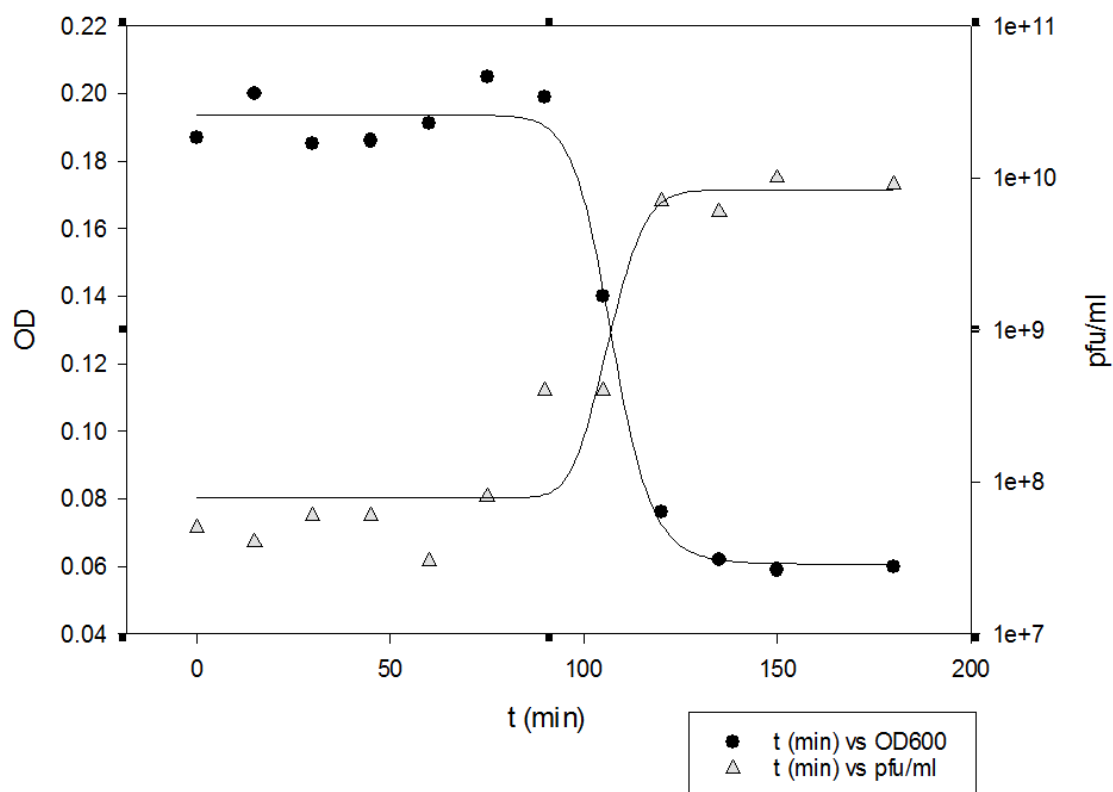


Figure 5.7 Correlation of *M. smegmatis* growth and D29 titre in infected cultures (MOI 10). A culture was infected at OD₆₀₀=0.2 (left Y-axis). The OD of the culture (black dots) was documented when culture samples were collected over 3 h. The phage titre (grey triangles) in culture supernatant was assessed by spot test and is given in pfu/ml (right Y-axis, logarithmic scale). Sudden onset of lysis was observed and sudden rise in phage titre were observed at around 110 min p.i.

5.3.5 Cell harvesting

In order to obtain intact infected cells and assure accuracy of time points, we tested two different methods of pelleting cells, centrifugation and filtration. Cells can remain viable during pelleting, hence also viral growth continues. Centrifugation was done at 4°C, 400 ml flask rotor FA 500, 10 000 × g. Unfortunately, pelleting of *M. smegmatis* requires over 30 min at high volumes. Vacuum filtration using 0.22 µm pore size Whatmann membrane allows for “pelleting” within 2-10 min. The culture is filtered and cells are retained on the filter. Cells can be scraped off the filter and immediately frozen at -80 °C. 200 ml can be filtered very fast (2-3 min), the speed then decreases drastically (300 ml takes 7-10 min). We chose to filter cells in aliquots of 300 mL for highest possible protein yield with fast filtration.

5.3.6 Phage protein pull-down via His-tagged mycobacterial RNA Polymerase

We tested various methods of cell lysis and salt concentrations for His-tag affinity chromatography. For lysis, BeadBeating was compared to sonication. Table 5.2 shows yield of purified RNAP from BeadBeating (1) and sonication (2). As comparable yield was achieved and sonication is faster and easier to use, we decided to continue using sonication.

M. smegmatis were sonicated in 7 ml ice cold grinding buffer with 20 µg/ml proteinase inhibitor at 20% amplitude, 15 sec pulses on/off, for 5 min. Care was taken to keep the lysates ice cold. Further, total yield of purified protein (RNAP) of late stage infected cells (sample 3) was found to be much lower, indicating that cells may become permeable during infection and suffer loss of protein during pelleting.

Table 5.2 Yield of purified His-tagged RNAP for different cell lysis methods. Protein was purified via affinity chromatography (Ni²⁺-NTA agarose binding to His-tag on RpoC). Starting material (200 mL early log phase cells) were lysed via BeadBeater (1) or sonication (2,3). Infected cells give lower yield, probably due to instability of cell membrane.

Sample	Method	Culture volume (ml, OD ₆₀₀ =0.2)	Total protein yield (µmol)
1 not infected	BeadBeater	200	0.156
2 not infected	Sonication	200	0.18
3 infected	Sonication	200	0.036

For pull down of phage proteins that interact with RNAP, different samples and controls were purified as described above (figure 5.3).

In contrast to use of His-Trap Ni-affinity chromatography as described in the methods section, here, Ni-affinity purification was performed using Ni²⁺-NTA agarose beads. The lysate in grinding buffer was centrifuged to remove cell debris (10 000 × g, 15 min, 4 °C), and filtered through 0.45 µm filter, 10 mM imidazole was added. 500 µL Ni²⁺-NTA agarose beads, previously equilibrated in 10 mM Imidazole Ni-column buffer, were applied to 7 ml cleared

lysate (= Load, see figure 5.3.5). Beads were incubated on ice with occasional shaking for about 15 min. Beads were allowed to settle by gravity and supernatant was removed (= flow through, FT). Beads were washed in 500 μ L wash buffer 3 times (= wash, W). For elution, beads were washed in 200 μ L elution buffer (100 mM imidazole) twice and then in 200 mM imidazole once (=eluates). After His-tag affinity purification via Ni^{2+} -NTA agarose beads we subjected different samples to denaturing SDS-polyacrylamide gel electrophoresis (SDS-PAGE) to detect RNAP as well as potential phage Gps. Figure 5.8 shows eluates from phage D29 (t = 15) infection which compared to control (uninfected) RNAP shows an additional band of ~55 kDa size. Coincidentally, wash fractions of both infected and not-infected purifications contain a 55 kDa band as well. This gel is representative of other gels from different rounds of infection. In most cases, no additional bands could be detected with the naked eye. Either way, the eluates from RNAP purification were pooled, concentrated and sent for MS analysis.

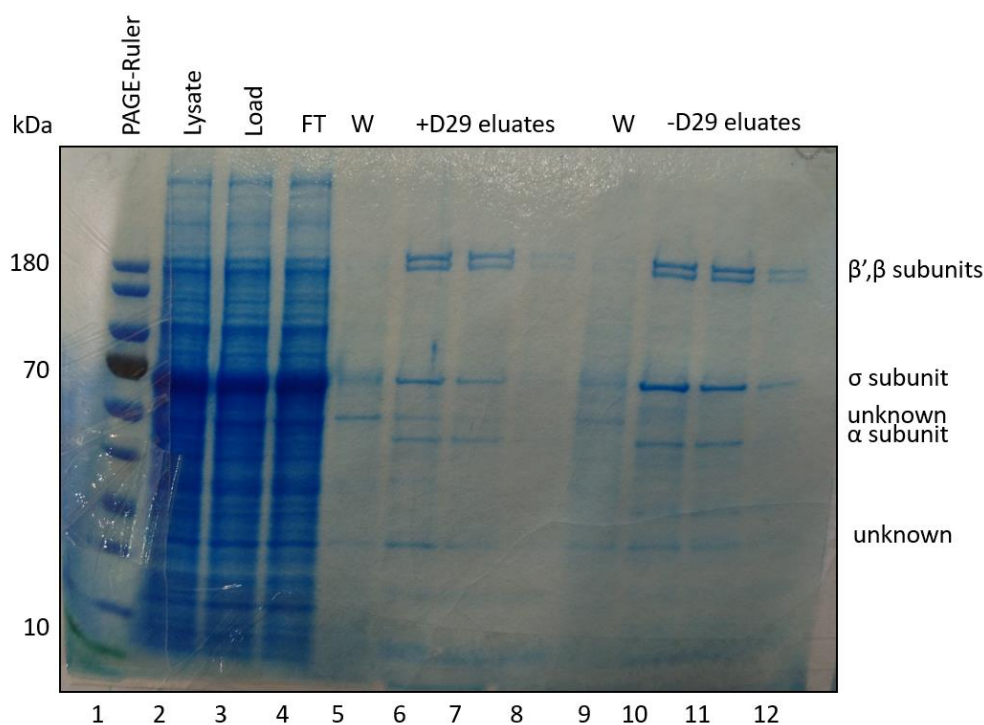


Figure 5.8 SDS-PAGE of RNAP pull-down from infected and uninfected cells. *M. smegmatis* strain SM07 (His-tagged RpoC) was infected (+D29) or not infected (-D29) with mycobacteriophage D29. At t=15 min, cultures were filtered, frozen, and then lysed by sonication (Lysate, lane 2). The lysate was cleared from cell debris (Load, lane 3), and incubated with Ni^{2+} -NTA agarose beads at 10 mM imidazole. The bead supernatant (FT, lane 4) was removed and bead were washed (W, lane 5 and 9) in 25 mM imidazole. Beads were eluted twice by 100 mM imidazole and once by 200 mM imidazole (lanes 6-8 and 10-12). 10 μ L samples of each step were loaded on this gel. Lanes 2-8 show samples from D29-infected cells, lanes 9-12 from not-infected cells (control).

5.3.7 Mass Spectrometry analysis of pull-down

Protein yield of infected cells was often very low with about 10-50 pmoles per purification. Thus pooled eluates were concentrated to smallest possible volume and loaded on SDS-PAGE. Electrophoresis was discontinued as soon as protein had completely entered the separating gel and was cut out. The gel slice was sent to Fingerprints Proteomics Facility at Dundee University for mass spectrometry. For each phage we analysed 5 samples: Infection of strain SM07, at two different times p.i. (15 and 40 min for D29), infection of wildtype strain at both time points, and strain SM07 without phage infection. From the results of infected pull-downs we subtracted all non-phage hits and Gps found in mock pull-down (see table 5.3). All found phage proteins were contained at 40 min p.i., only Gp68 was also found at 15 min p.i. No phage protein was found in negative control.

Table 5.3 Mycobacteriophage Gps found in samples of infected RNAP purifications. MS was performed by Fingerprints Proteomics Facility at the University of Dundee. Five Gps encoded by phage D29 were found only in infected SM07, not in wildtype or negative control (no infection). Four of these are previously uncharacterised proteins. (Gp53, Gp55, Gp65, and Gp68). Protein size in amino acid (AA), and molecular weight (MW) are given as well as calculated isoelectric point (pI).

Protein name and accession	pulled down via RpoC t=15 min	pulled down via RpoC t=40 min	MW [kDa]	pI
Probable thymidylate synthase OS= <i>Mycobacterium</i> phage D29 GN=48 PE=3 SV=1 - [THYX_BPMD2]	no	yes	26.5	5.17
Gene 53 protein OS= <i>Mycobacterium</i> phage D29 GN=53 PE=4 SV=1 - [VG53_BPMD2]	no	yes	26.3	7.20
Gene 55 protein OS= <i>Mycobacterium</i> phage D29 GN=55 PE=4 SV=1 - [VG55_BPMD2]	no	yes	17.3	5.41
Gene 65 protein OS= <i>Mycobacterium</i> phage D29 GN=65 PE=4 SV=1 - [VG65_BPMD2]	no	yes	25.1	4.65
Gene 68 protein OS= <i>Mycobacterium</i> phage D29 GN=68 PE=4 SV=1 - [VG68_BPMD2]	yes	yes	8.7	10.27

5.3.8 Cloning of phage proteins

In order to identify transcription inhibitors, we needed to test potential candidates individually. For this, the phage proteins identified in mass spectrometry which were not found in controls and had not been previously identified (D29 proteins Gp53, Gp55, Gp65 and Gp68) were PCR amplified with custom primers from IDT and cloned into T7 expression system vector pET28a for recombinant expression in *E. coli*, as described in methods. Not all phage proteins could be purified via this route due to problems at cloning or overexpression. If a protein is cytotoxic in the overexpression system (*E. coli*), only mutants will be overexpressed (e.g. which developed a frameshift mutation in the gene). If too much protein is overexpressed, misfolding can occur and inclusion bodies can form. These need to be purified at denaturing conditions, and renaturation is not always possible. Gp65 was not overexpressed. The reason of this is uncertain. Of this D29 set, Gp53, Gp55 and Gp68 have been tested in *in vitro* transcription. As only Gp53 showed distinct inhibition of *M. smegmatis* RNAP, further graphs will focus on Gp53 results.

5.3.9 Purification of phage proteins

As phage proteins were previously uncharacterised, we ran a preliminary test for optimal overexpression conditions, considering different temperatures and durations of induction. Cells were grown in selective medium to $OD_{600} = 0.5$ at 37 °C. After temperature shift to 30 °C or room temperature (ca 20-25°C), cells were induced with 1 mM IPTG. After 3 h of overexpression, samples were collected, lysed by heat and cell debris was removed by centrifugation. SDS-PAGE of lysates showed that Gp53 formed inclusion bodies if overexpressed at 30 °C, thus we decided to overexpress at room temperature for 3 h. A larger culture (2 L) was inoculated in complex medium with antibiotic kanamycin for pET28a overexpression vector. Colder temperatures lead to a slower protein production which can be beneficial to prevent formation of inclusion bodies. Overexpression was induced 20 min after transferring the culture to room temperature, at $OD_{600} = 0.5$, and harvested cells 3 h later, at $OD_{600} = 0.95$.

Cells were lysed and Gp53 was purified via Ni^{2+} -affinity chromatography as described in methods, in presence of 300 mM NaCl. After a first purification, bands that resemble RNAP subunits β , β' and σ^{70} were present as contaminants in Gp53 eluates. We cannot rule out that EcRNAP was co-purified with Gp53. Since relative concentration of Gp53 according to SDS-

PAGE was much higher than protein contaminants, we proceeded to investigate this Gp53 purification in initial tests (section 5.3.8).

However, as initial results were promising, Gp53 was repurified at higher NaCl concentration (600 mM), and using ion exchange to eliminate contaminants (figure 5.9).

Further, the N-terminal His-tag of Gp53 was removed by Thrombin cleavage. This was done in case Gp53 activity or structure depends on a N-terminal domain. A thrombin cleavage site is positioned on the protein between His-tag and start codon. Gp53 purification shown in lane 7, figure 5.9 was subjected to thrombin cleavage. The protein is incubated with Thrombin enzyme overnight at 4 °C (according to Thrombin kit protocol). Thrombin was removed via affinity resin. Residual His-tagged Gp53 was removed by Ni²⁺-NTA agarose beads.

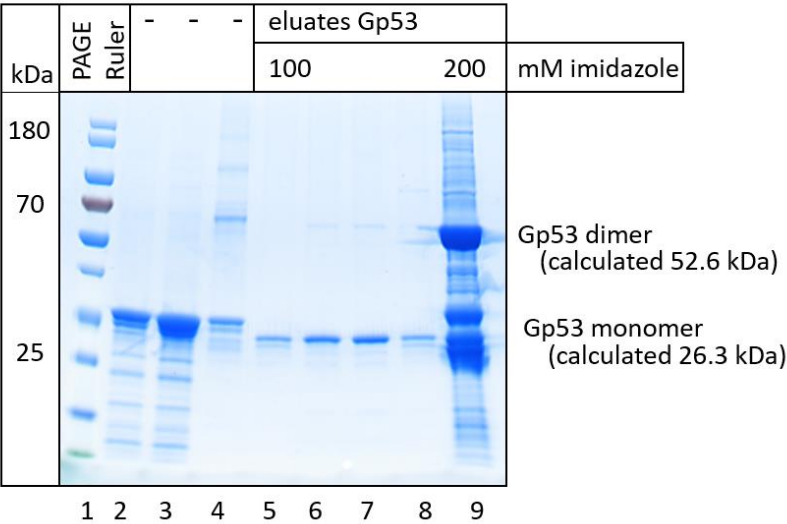


Figure 5.9 Purification of His-Gp53. Lanes 2-4 can be disregarded as they show the purification of a different protein (NudC). His-Gp53 was purified by Ni²⁺-affinity chromatography and then loaded onto an ion exchange column (Resource Q). Resource Q buffers without EDTA and DTT were used. Lanes 5-9 show different purifications pooled after ion exchange chromatography of Gp53 in 100 mM imidazole, and after dialysis in storage buffer containing DTT. Lane 5 contained Gp53 eluted as monomer from ion exchange, 6 - 8 were eluted as monomer and homodimer.

5.3.10 His-tagged Gp53 supports transcription of long products

To assess the effect of purified phage proteins on transcription, it is added to *in vitro* transcription assay. 1 pmole RNAP with 2 pmoles SigA was mixed with 3 pmoles of template DNA *rrsA*. This mocobacterial promoter precedes the gene of 16S rRNA (Dal Molin *et al.*, 2018). 3 pomoles Gp53 were added to half of the assays (ratio RNAP:Gp53 1:3). We also tested Gp53 effect in presence of *M. smegmatis* transcription factor RbpA (1:3 RNAP to

RbpA). Transcription assays further contained 100 μM of each NTP except UTP (10 μM) which was additionally provided as $\alpha(\text{P}^{32})$ -UTP for radioactive labelling. This allows for “run-off” products, where RNAP transcribes from promoter to the end of the linear DNA, or until a pause/termination signal is encountered. The *M. smegmatis* transcription buffer containing 10 mM MgCl was added last to start the reaction. After 10 min incubation at 37 $^{\circ}\text{C}$, reactions were stopped with transcription stop buffer and then loaded on 15% polyacrylamide gel.

Gp53 increases transcription by more than two fold without transcription activator RbpA (figure 5.10). When RbpA was present, transcription was 5% stronger upon addition of Gp53. This does not mean that Gp53 replaces transcription activation by RbpA. It might act as transcriptional activator, or act as anti-terminator. Contamination by EcRNAP might temper with the results. Therefore, after this initial experiment, Gp53 was repurified. If Gp53 co-purifies EcRNAP at medium salt concentrations (300 mM NaCl) this could be an indication that Gp53 interacts with bacterial RNAP directly, and the binding region might be an evolutionarily conserved region of RNAP. Alternatively, it might co-purify RNAP because of a strong interaction of both Gp53 and RNAP with promoter DNA.

5.3.11 Gp53-mediated transcription inhibition relies on free N-terminal domain

We removed the N-terminal His-tag from Gp53 using the Thrombin cleavage kit. We performed a run-off transcription assay on *M. smegmatis* promoter *rrsA* as described above and observed that Gp53 inhibits transcription by MsRNAP holoenzyme, with increased effect in presence of transcription activator RbpA (figure 5.11 B). Addition of His-Gp53 to MsRNAP holoenzyme reduced transcription by 17%. Addition of free Gp53 to holoenzyme resulted in loss of transcription efficiency by 70 %. Addition of RbpA alone to holoenzyme did not enhance transcription on *rrsA*, but addition of His-Gp53 to MsRNAP- σ^A -RbpA increased transcription by 30%. This is another indicator that the previous transcription enhancement by Gp53 (section 5.3.9) is due to an activator activity of Gp53. This activator activity might rely on misfolding due to the N-terminal His-tag. Possibly, Gp53 has several roles. Addition of non-tagged Gp53 to RNAP- σ^A -RbpA decreased transcription efficiency by 87.7% as compared to RNAP- σ^A -RbpA alone and by 90% as compared to RNAP- σ^A -RbpA-His-Gp53. This indicates that Gp53 N-terminal domain is likely involved in a) protein folding or complex formation or b) inhibitory activity (via RNAP/RbpA binding, DNA/RNA binding or enzymatic activity). Seemingly, involvement of RbpA increases inhibition by free Gp53. The effect might thus be promoter-dependent. RbpA was previously shown to confer

MsRNAP holoenzyme stability on promoters lacking -35 and extended -10 elements. *rrsA* has -35 and a regular -10 element (Perumal *et al.*, 2018).

We tested the effect of Gp53 on transcription from promoter P_left, a mycobacteriophage promoter (see figure 5.11 A and C). P_left is the early lytic promoter of lysogenic mycobacteriophage L5, a close relative of D29 (Brown *et al.*, 1997). This promoter has both a -35 and an extended -10 element, thus in theory it does not require RbpA. We observed that Gp53 inhibits transcription from P_left by 58% as compared to transcription by MsRNAP holoenzyme, and by 88% in presence of RbpA. At this point we cannot make any detailed conclusions about the mechanism of inhibition. Interaction of Gp53 with RbpA seems likely.

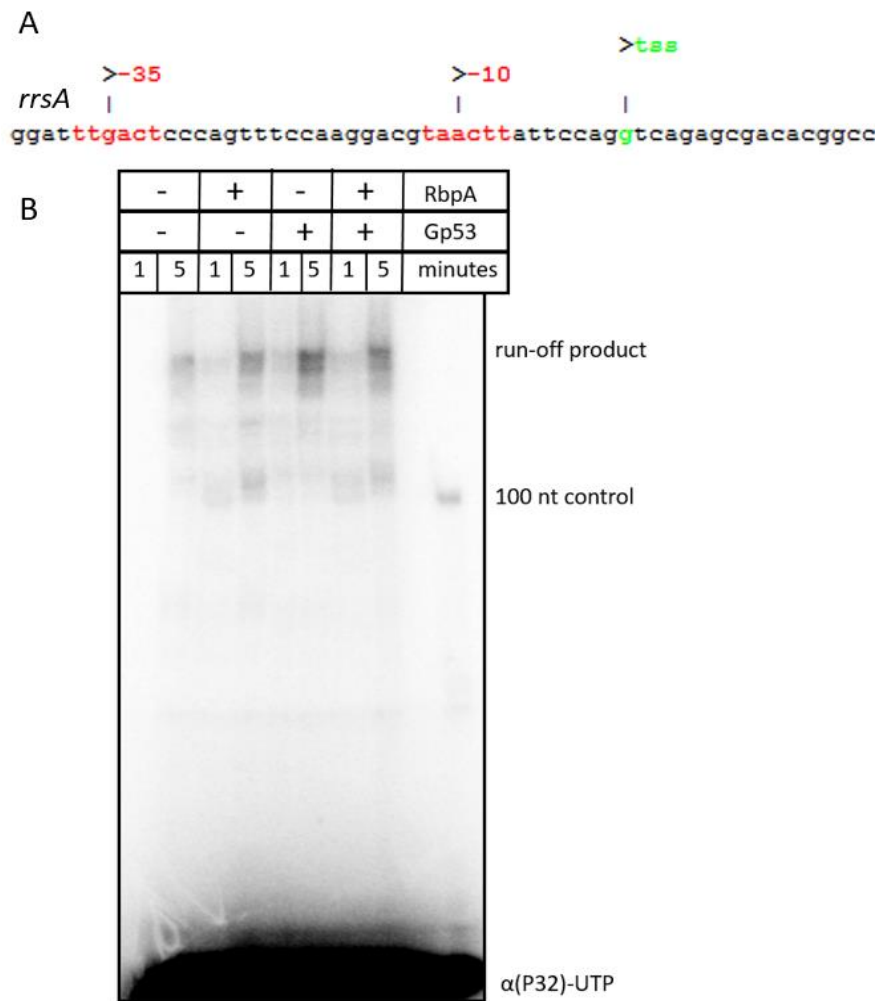


Figure 5.10 Transcription activation by first Gp53 purification. A. promoter region of the *rrsA* template. Promoter region -35 and -10 are marked in red and transcription start site is marked in green. The full template allows for run-off products of up to 400 nt. The run-off products are likely due to a termination signal on the sequence. His-tagged Gp53 enhanced transcription on promoter *rrsA* in presence and in absence of RbpA.

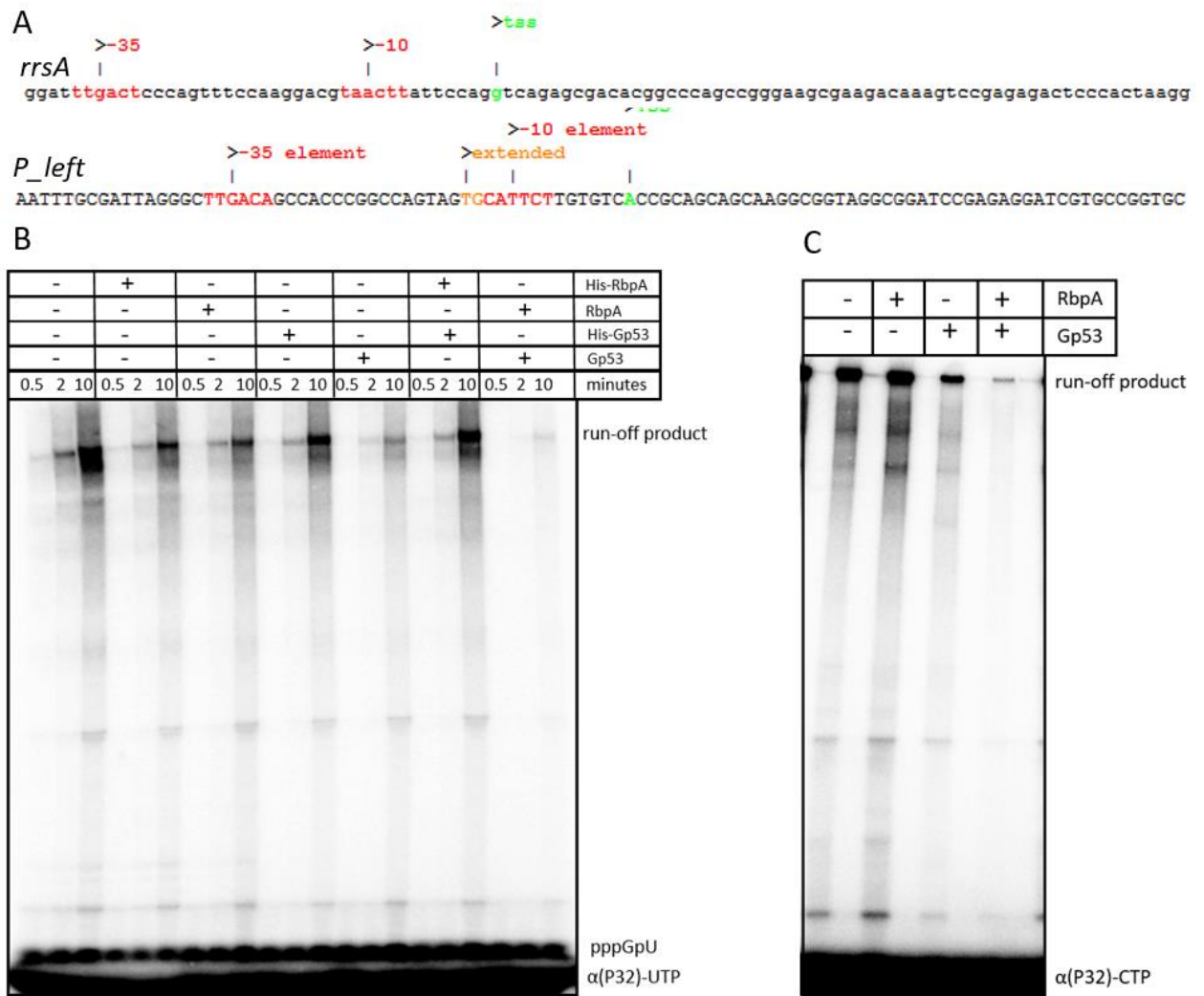


Figure 5.11 Gp53 inhibits transcription on two promoters. A. Promoters *rrsA* and *P_left* with σ -binding elements (red,orange) and TSS (green). B. Transcription of MsRNAP on *rrsA* in presence/absence of RbpA and Gp53. His-tagged and non-tagged versions of both RbpA and Gp53 were tested. Non-tagged Gp53 visibly inhibits transcription in absence of RbpA, but even more so in presence of RbpA. His-tagged Gp53 seems to act as weak transcriptional activator C. Transcription of MsRNAP on promoter *P_left* in presence/ absence of RbpA and non-tagged Gp53. Gp53 reduces transcription efficiency in absence but more so in presence of RbpA.

5.3.12 Gp53 inhibits *M. smegmatis* RNAP, but not *E. coli* RNAP

One of the advantages of enzybiotics is that the narrow host range of bacteriophages is generally reflected in their narrow specificity to orthologous proteins. We therefore tested, if Gp53 affects *E. coli* transcription *in vitro*. We designed a modified version of promoter *P_left* to facilitate single nucleotide resolution observations. The reactions were performed as described earlier with the modification that only ATP, CTP and GTP substrates were added). EcRNAP transcription was much more efficient on *P_left* than *M. smegmatis* RNAP

transcription first (before RNAP and DNA, lane 3) or last (after RNAP-promoter complex formation, lane 4) does not affect EcRNAP transcription efficiency. In contrast, addition of Gp53 to *M. smegmatis* *in vitro* transcription resulted in reduced efficiency of RNAP in any case (lanes 8-11), but inhibition was more prominent, when Gp53 was added to RNAP before DNA was added (lanes 8 and 9). We conclude that Gp53 binding to RNAP holoenzyme (in presence and absence of RbpA) reduces transcription, and this binding might or might not hinder RNAP association to the promoter. We can further conclude that Gp53 does not inhibit transcription via blocking of the RNAP-binding site on the promoter.

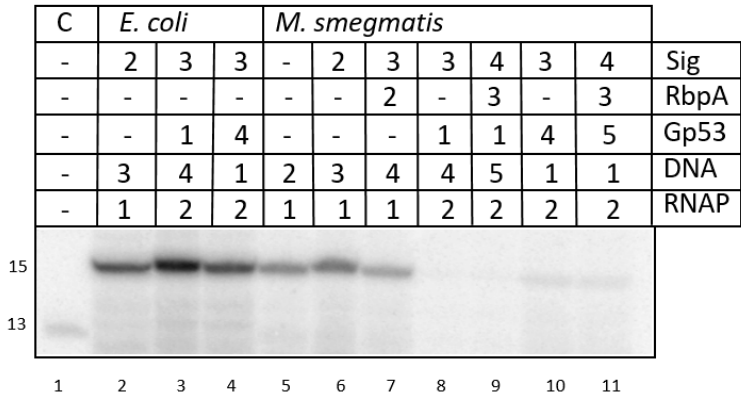


Figure 5.13. Gp53 inhibits transcription via protein-binding. Transcription components were added in different chronological orders to determine whether Gp53 inhibits transcription via unspecific binding of promoter DNA (P_{left} mod), or if inhibition relies on interaction with RNAP after RNAP-promoter complexation. *E. coli* (lanes 1-3) or *M. smegmatis* (lanes 5-11) RNAP was added before or after DNA and Gp53. It becomes visible that Gp53 inhibits transcription if added before and after RNAP-promoter complex formation, but inhibition is most efficient when GP53 is added to RNAP before DNA. A 13mer transcript was loaded as size control (C).

5.3.14 Further improvement of the protocol

This protocol was shown to allow for isolation of phage proteins that inhibit transcription via interaction with RNAP. However, some improvements can still be made.

In order to omit the step of His-tag removal by Thrombin cleavage, proteins can be cloned and purified with both C-terminal and N-terminal tags (two versions of each protein).

In several mass spectrophotometric analyses, structural proteins were found. This could be an indicator that time points for harvesting cells were chosen too late, such that virion assembly had already commenced. At the point of virion assembly, the phage might have ceased its manipulation of host RNAP. However, structural genes can affect host transcription as well as shown for *E. coli* phage Gp Psu, a capsid protein which also acts as anti-terminator for Rho-

dependent terminator (Linderoth and Calendar, 1991; Ghosh *et al.*, 2018), thus they should not be prematurely excluded from further analysis.

Further, purified Gps were not always active in *in vitro* transcription. This likely means that the protein is not involved in transcription regulation. However, this protocol does not take into consideration potential requirement for coenzymes such as metal ions. If a phage protein requires a ligand that is not abundant in the complex medium it might not be active or stable after purification. Although this limitation might cause us to overlook some interesting phage transcription regulators, it would be too much work to consider all potential coenzymes for overexpression. Instead, we could reduce loss of metal coenzymes by omitting EDTA in purification buffers.

5.4 Conclusion

We have developed a protocol that allows for screening for phage regulators of mycobacterial transcription via interaction with RNAP or accessory proteins. This method has thus far been applied for mycobacteriophage D29, TM4 and Adephagia mutant $\Delta 41\Delta 43$. Only one species-specific inhibitor was found. Gp53 from mycobacteriophage D29 is here shown to inhibit *M. smegmatis* transcription *in vitro* on two different promoters, *rrsA* and P_{left}. Both promoters have -35 and -10 elements and bind the housekeeping σ^A . Promoter *rrsA* is an rRNA promoter. These are often regulated independently of other promoters, by signalling of metabolic state and several feedback mechanisms (Jacob, 1995; Burgos *et al.*, 2017).

Promoter P_{left} is a mycobacteriophage L5 early lytic promoter which is regulated by the phage repressor (Brown *et al.*, 1997). It is possibly an extended -10 promoter, as it contains a TG motif upstream of the -10 element. Phage D29 is a lytic phage in both *M. smegmatis* and *M. tuberculosis*. It is very closely related to temperate phage L5 (Hatfull, 2010b). Both are Cluster A2 mycobacteriophages. They differ by deletion of a 3.6 kb deletion which includes loss of repressor gene Gp71. However, phage attachment site as well as integrase gene are present on the D29 genome, and it is affected by superinfection immunity by L5.

Superinfection immunity is conferred by repressor proteins which bind to certain sites on the phage genome and repress expression of genes involved in lytic lifecycle. The repressor of one phage can suppress gene expression of a secondary infecting phage if repressor binding sites comply. Thus a bacterium that is infected with one phage can be immune to a second phage (Donnelly-Wu, Jacobs and Hatfull, 1993; Ford, Sarkis, *et al.*, 1998). L5 operators have

been found to overlap with a variety of L5 promoters, including the early lytic promoter P_{left} which was used in this study (Ford, Sarkis, *et al.*, 1998). P_{left} is also present on D29 genome but has some nucleotide substitutions as compared to P_{left} from L5, mainly within the -10 element (Ford, Sarkis, *et al.*, 1998). L5 also encodes a homologue of Gp53 (82% sequence identity to D29 Gp53). The function of this protein has not been identified. As Gp53 inhibits transcription on both promoters we can speculate that its' inhibition does not specifically target P_{left} and it is not a phage repressor. Instead, Gp53 might function as a general transcription regulator.

Interestingly, we have further observed a likely interaction between Gp53 and host-encoded transcription activator RbpA. RbpA is essential in mycobacteria (Forti *et al.*, 2011; Hu *et al.*, 2012; Tabib-Salazar *et al.*, 2013; Hubin *et al.*, 2015; Berger *et al.*, 2017). RbpA C-terminus binds to region 2 of the housekeeping sigma, but not alternative sigma (Tabib-Salazar *et al.*, 2013). RbpA- σ^A binds tightly to -10 promoter element and stabilises open complex formation (Elizabeth A Hubin *et al.*, 2017) (see figure 5.14). In *M. tuberculosis*, 73% of promoters have a -10 element of consensus sequence TANNNT, 7 of which are preceded with an extended TGN. -35 elements are rare in the mycobacterial genome (Cortes *et al.*, 2013). Thus, Gp53 inhibition might be predominantly directed towards σ^A –promoters, via interaction with RbpA. This mode of transcription inhibition might allow the phage to shut off the majority of host transcription without completely inhibition host RNAP. This hypothesis requires further investigation.

Further, we have observed that Gp53 activity likely relies on its N-terminus, as N-terminally His-tagged Gp53 does not inhibit transcription, but might acts as anti-terminator or transcription activator *in vitro*. The protein might have several functional roles. Gps which act on several stages of transcription have been identified in other phage-host systems as well, such as Gp39 of *T. thermophilus* phage P23-45, which acts both in σ -appropriation as well as anti-termination.

5.5 Outlook/ Perspectives

The outlined methodology has been shown to facilitate rapid identification of phage-encoded transcription inhibitors. The steps to be taken are i) Choice of appropriate time points for harvesting infected cells and controls (non-tagged RNAP and non-infected cells). ii) Cell lysis and purification of *M. smegmatis* RNAP via His-tag. iii) Mass spectrometric analysis of

proteins iv) Cloning and overexpression of phage proteins which are not contained in controls
vi) Further *in vitro* analysis of individual proteins. We found an inhibitor of mycobacterial transcription, Gp53 from phage D29. Extensive *in vitro* analysis and characterisation was not within the scope of this thesis. Further experiments need to identify i) binding partner(s) of Gp53 (protein-protein interactions as well as protein-nucleic acid interactions) ii) the precise step of transcription it interferes with and iii) its' precise mode of action. A further essential experiment would be effectivity on *M. tuberculosis* transcription.

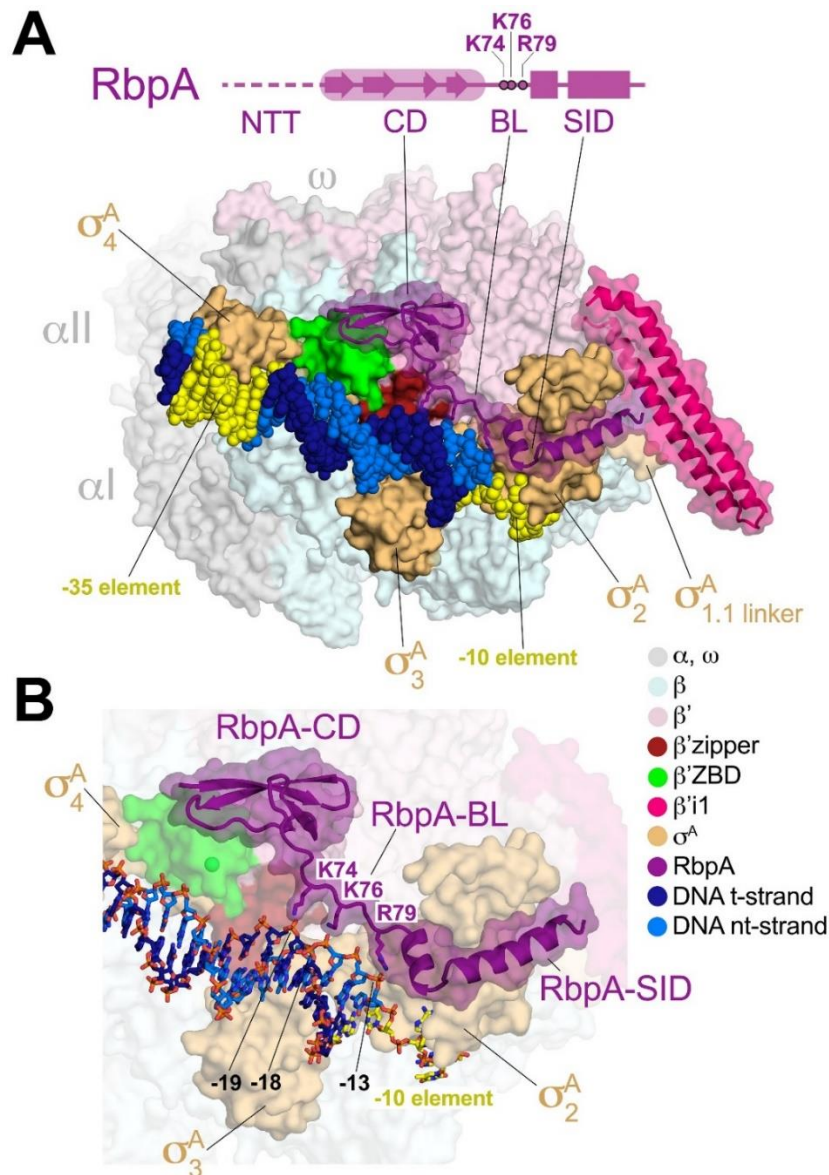


Figure 5.14 Structure of mycobacterial transcription factor RbpA in complex with RNAP- σ^A holoenzyme and DNA. From (Elizabeth A. Hubin *et al.*, 2017). A. Structural domains of RbpA are N-terminal tail (NTT), core domain (CD), basic linker (BL) and the C-terminal sigma-interacting domain (SID). B. CD contacts β' Zn^{2+} -binding domain (β' ZBD) and β' zipper, both belonging to conserved RNAP “clamp”. BL interacts with DNA and SID with σ^A region 2.

6 References

- Adebali, O. *et al.* (2017) 'Genome-wide transcription-coupled repair in *Escherichia coli* is mediated by the Mfd translocase.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 114(11), pp. E2116–E2125. doi: 10.1073/pnas.1700230114.
- Adelman, K. *et al.* (2004) 'Molecular Mechanism of Transcription Inhibition by Peptide Antibiotic Microcin J25', *Molecular Cell*. Cell Press, 14(6), pp. 753–762. doi: 10.1016/J.MOLCEL.2004.05.017.
- Akabayov, B. *et al.* (2009) 'DNA Recognition by the DNA Primase of Bacteriophage T7: A Structure–Function Study of the Zinc-Binding Domain [†]', *Biochemistry*, 48(8), pp. 1763–1773. doi: 10.1021/bi802123t.
- Anton, V., Rougé, P. and Daffé, M. (1996) 'Identification of the sugars involved in mycobacterial cell aggregation', *FEMS Microbiology Letters*. John Wiley & Sons, Ltd (10.1111), 144(2–3), pp. 167–170. doi: 10.1016/0378-1097(96)00356-4.
- Aravind, L., Leipe, D. D. and Koonin, E. V (1998) 'Toprim--a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins', *Nucleic Acids Research*, 26(18), pp. 4205–4213. doi: 10.1093/nar/26.18.4205.
- Arias-Palomo, E. *et al.* (2013) 'The Bacterial DnaC Helicase Loader Is a DnaB Ring Breaker', *Cell*, 153(2), pp. 438–448. doi: 10.1016/j.cell.2013.03.006.
- Arjanova, O. V *et al.* (2011) 'Adjunct oral immunotherapy in patients with re-treated, multidrug-resistant or HIV-coinfected TB', *Immunotherapy*. Future Medicine Ltd London, UK , 3(2), pp. 181–191. doi: 10.2217/imt.10.96.
- Artsimovitch, I. and Landick, R. (2000) 'Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals', *Proceedings of the National Academy of Sciences*, 97(13), pp. 7090–7095. doi: 10.1073/pnas.97.13.7090.
- Augustin, M. A., Huber, R. and Kaiser, J. T. (2001) 'Crystal structure of a DNA-dependent RNA polymerase (DNA primase).', *Nature Structural Biology*, 8(1), pp. 57–61. doi: 10.1038/83060.
- Bae, B. *et al.* (2015) 'Structure of a bacterial RNA polymerase holoenzyme open promoter complex', *eLife*, 4. doi: 10.7554/eLife.08504.
- Bae, S.-H. *et al.* (2001) 'RPA governs endonuclease switching during processing of Okazaki fragments in eukaryotes', *Nature*, 412(6845), pp. 456–461. doi: 10.1038/35086609.
- Bailey, S., Wing, R. A. and Steitz, T. A. (2006) 'The Structure of *T. aquaticus* DNA Polymerase III Is Distinct from Eukaryotic Replicative DNA Polymerases', *Cell*, 126(5), pp. 893–904. doi: 10.1016/j.cell.2006.07.027.
- Balakrishnan, L. and Bambara, R. A. (2013) 'Okazaki fragment metabolism.', *Cold Spring Harbor perspectives in biology*. Cold Spring Harbor Laboratory Press, 5(2), p. a010173. doi: 10.1101/cshperspect.a010173.
- Balakrishnan, L. and Bambara, R. A. (2016) 'Okazaki Fragment Metabolism', *Cold Spring Harbor Perspectives in Biology*, 5, pp. 1–12. doi: 10.1101/cshperspect.a010173.
- Bambara, R. A., Murante, R. S. and Henricksen, L. A. (1997) 'Enzymes and reactions at the eukaryotic DNA replication fork.', *The Journal of biological chemistry*. American Society for

- Biochemistry and Molecular Biology, 272(8), pp. 4647–50. doi: 10.1074/jbc.272.8.4647.
- Bandyra, K. J. *et al.* (2013) ‘The social fabric of the RNA degradosome’, *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*. Elsevier, 1829(6–7), pp. 514–522. doi: 10.1016/J.BBAGRM.2013.02.011.
- Baranovskiy, A. G. *et al.* (2015) ‘Crystal structure of the human primase’, *Journal of Biological Chemistry*, 290(9), pp. 5635–5646. doi: 10.1074/jbc.M114.624742.
- Baranovskiy, A. G. *et al.* (2016) ‘Insight into the Human DNA Primase Interaction with Template-Primer.’, *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 291(9), pp. 4793–802. doi: 10.1074/jbc.M115.704064.
- Barchiesi, A. and Vascotto, C. (2019) ‘Transcription, processing, and decay of mitochondrial RNA in health and disease’, *International Journal of Molecular Sciences*, 20(9). doi: 10.3390/ijms20092221.
- Basu, R. S. *et al.* (2014) ‘Structural Basis of Transcription Initiation by Bacterial RNA Polymerase Holoenzyme’, *Journal of Biological Chemistry*, 289(35), pp. 24549–24559. doi: 10.1074/jbc.M114.584037.
- Batada, N. N. *et al.* (2004) ‘Diffusion of nucleoside triphosphates and role of the entry site to the RNA polymerase II active center’, *Proceedings of the National Academy of Sciences*, 101(50), pp. 17361–17364. doi: 10.1073/pnas.0408168101.
- Batdelger, D. *et al.* (2008) ‘Open-label trial of therapeutic immunization with oral V-5 Immunitor (V5) vaccine in patients with chronic hepatitis C’, *Vaccine*. Elsevier, 26(22), pp. 2733–2737. doi: 10.1016/J.VACCINE.2008.03.021.
- Batey, R. T., Rambo, R. P. and Doudna, J. A. (1999) ‘Tertiary Motifs in RNA Structure and Folding’, *Angewandte Chemie International Edition*. John Wiley & Sons, Ltd, 38(16), pp. 2326–2343. doi: 10.1002/(SICI)1521-3773(19990816)38:16<2326::AID-ANIE2326>3.0.CO;2-3.
- Beelman, C. A. and Parker, R. (1995) ‘Degradation of mRNA in eukaryotes.’, *Cell*. Elsevier, 81(2), pp. 179–83. doi: 10.1016/0092-8674(95)90326-7.
- Beinert, H., Holm, R. H. and Münck, E. (1997) ‘Iron-Sulfur Clusters: Nature’s Modular, Multipurpose Structures’, *Science*, 277(5326), pp. 653–659. doi: 10.1126/science.277.5326.653.
- Bell, S. D. (2019) ‘Initiating DNA replication: a matter of prime importance.’, *Biochemical Society transactions*. Portland Press Limited, 47(1), pp. 351–356. doi: 10.1042/BST20180627.
- Bennett, B., Kimball, E. and Gao, M. (2009) ‘Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*.’, *Nature Chemical Biology*, 5(8), pp. 593–599. doi: 10.1038/nchembio.186.Absolute.
- Berdygulova, Z. *et al.* (2011) ‘Temporal regulation of gene expression of the *Thermus thermophilus* bacteriophage P23-45.’, *Journal of molecular biology*. NIH Public Access, 405(1), pp. 125–42. doi: 10.1016/j.jmb.2010.10.049.
- Berdygulova, Z. *et al.* (2012) ‘A novel phage-encoded transcription antiterminator acts by suppressing bacterial RNA polymerase pausing’, *Nucleic Acids Research*. Oxford University Press, 40(9), pp. 4052–4063. doi: 10.1093/nar/gkr1285.

- Berger, J. M. *et al.* (2017) 'Structure and function of the mycobacterial transcription initiation complex with the essential regulator RbpA', *eLife*, 6, pp. 1–40. doi: 10.7554/eLife.22520.001.
- Berkower, I., Leis, J. and Hurwitz, J. (1973) 'Isolation and characterization of an endonuclease from *Escherichia coli* specific for ribonucleic acid in ribonucleic acid-deoxyribonucleic acid hybrid structures.', *The Journal of biological chemistry*, 248(17), pp. 5914–21. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/4580046> (Accessed: 2 October 2019).
- Bertozzi Silva, J., Storms, Z. and Sauvageau, D. (2016) 'Host receptors for bacteriophage adsorption', *FEMS Microbiology Letters*. Edited by A. Millard, 363(4), p. fnw002. doi: 10.1093/femsle/fnw002.
- Bessman, M. J. *et al.* (2001) 'The Gene *ygdP*, Associated with the Invasiveness of *Escherichia coli* K1, Designates a Nudix Hydrolase, Orf176, Active on Adenosine (5')-Pentaphospho-(5')-adenosine (Ap5A)', *Journal of Biological Chemistry*, 276(41), pp. 37834–37838. doi: 10.1074/jbc.M107032200.
- Bhattacharyya, S. and Griep, M. A. (2000) 'DnaB Helicase Affects the Initiation Specificity of *Escherichia coli* Primase on Single-Stranded DNA Templates [†]', *Biochemistry*, 39(4), pp. 745–752. doi: 10.1021/bi991555d.
- Bird, J. G. *et al.* (2016) 'The mechanism of RNA 5' capping with NAD⁺, NADH and desphospho-CoA', *Nature*, 535(7612), pp. 444–447. doi: 10.1038/nature18622.
- Bird, J. G., Basu, U., Kuster, D., Ramachandran, A., Grudzien-Nogalska, E., Towheed, A., *et al.* (2018) 'Highly efficient 5' capping of mitochondrial RNA with NAD⁺ and NADH by yeast and human mitochondrial RNA polymerase', *eLife*, 7. doi: 10.7554/eLife.42179.
- Bird, J. G., Basu, U., Kuster, D., Ramachandran, A., Grudzien-Nogalska, E., Kiledjian, M., *et al.* (2018) 'Mitochondrial RNA capping: highly efficient 5'-RNA capping with NAD and NADH by yeast and human mitochondrial RNA polymerase', *bioRxiv*, 381160, p. manuscript. doi: 10.1101/381160.
- Blumberg, H. M. *et al.* (2003) 'American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America', *American Journal of Respiratory and Critical Care Medicine*, 167(4), pp. 603–662. doi: 10.1164/rccm.167.4.603.
- Bochkareva, A. *et al.* (2012) 'Factor-independent transcription pausing caused by recognition of the RNA-DNA hybrid sequence.', *The EMBO journal*. European Molecular Biology Organization, 31(3), pp. 630–9. doi: 10.1038/emboj.2011.432.
- Borukhov, S. and Nudler, E. (2008) 'RNA polymerase: the vehicle of transcription', *Trends in Microbiology*. Elsevier Current Trends, 16(3), pp. 126–134. doi: 10.1016/J.TIM.2007.12.006.
- Van den Bossche, A. *et al.* (2016) 'Structural elucidation of a novel mechanism for the bacteriophage-based inhibition of the RNA degradosome', *eLife*, 5. doi: 10.7554/eLife.16413.
- Botella, L. *et al.* (2017) 'Depleting *Mycobacterium tuberculosis* of the transcription termination factor Rho causes pervasive transcription and rapid death.', *Nature communications*. Nature Publishing Group, 8, p. 14731. doi: 10.1038/ncomms14731.
- Boudet, J. *et al.* (2015) 'Structures to complement the archaeo-eukaryotic primases catalytic cycle description: What's next?', *Computational and Structural Biotechnology Journal*. Elsevier B.V., 13, pp. 339–351. doi: 10.1016/j.csbj.2015.04.006.

- Bowman, G. D., O'Donnell, M. and Kuriyan, J. (2004) 'Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex', *Nature*. Nature Publishing Group, 429(6993), pp. 724–730. doi: 10.1038/nature02585.
- Brandis, G. *et al.* (2012) 'Fitness-compensatory mutations in rifampicin-resistant RNA polymerase', *Molecular Microbiology*. John Wiley & Sons, Ltd (10.1111), 85(1), pp. 142–151. doi: 10.1111/j.1365-2958.2012.08099.x.
- Brandis, G. *et al.* (2015) 'Comprehensive phenotypic characterization of rifampicin resistance mutations in *Salmonella* provides insight into the evolution of resistance in *Mycobacterium tuberculosis*', *Journal of Antimicrobial Chemotherapy*, 70(3), pp. 680–685. doi: 10.1093/jac/dku434.
- Brennan, M. J. (2017) 'Biofilms and *Mycobacterium tuberculosis*.' *Infection and immunity*. American Society for Microbiology Journals, 85(10), pp. e00411-17. doi: 10.1128/IAI.00411-17.
- Brown, K. L. *et al.* (1997) 'Transcriptional silencing by the mycobacteriophage L5 repressor', *EMBO Journal*. doi: 10.1093/emboj/16.19.5914.
- Broxmeyer, L. *et al.* (2002) 'Killing of *Mycobacterium avium* and *Mycobacterium tuberculosis* by a Mycobacteriophage Delivered by a Nonvirulent Mycobacterium: A Model for Phage Therapy of Intracellular Bacterial Pathogens', *The Journal of Infectious Diseases*. Oxford University Press, 186(8), pp. 1155–1160. doi: 10.1086/343812.
- Budd, M. E. *et al.* (2011) 'Inviability of a *DNA2* deletion mutant is due to the DNA damage checkpoint', *Cell Cycle*, 10(10), pp. 1690–1698. doi: 10.4161/cc.10.10.15643.
- Burgos, H. L. *et al.* (2017) 'Roles of Transcriptional and Translational Control Mechanisms in Regulation of Ribosomal Protein Synthesis in *Escherichia coli*.' *Journal of bacteriology*. American Society for Microbiology Journals, 199(21), pp. e00407-17. doi: 10.1128/JB.00407-17.
- Buskily, A. A. A., Kannaiah, S. and Amster-Choder, O. (2014) 'RNA localization in bacteria', *RNA Biology*, pp. 1051–1060. doi: 10.4161/rna.36135.
- Bustamante, N. *et al.* (2010) 'Cpl-7, a lysozyme encoded by a pneumococcal bacteriophage with a novel cell wall-binding motif', *Journal of Biological Chemistry*, 285(43), pp. 33184–33196. doi: 10.1074/jbc.M110.154559.
- Butcher, S. E. and Pyle, A. M. (2011) 'The Molecular Interactions That Stabilize RNA Tertiary Structure: RNA Motifs, Patterns, and Networks', *Accounts of Chemical Research*. American Chemical Society, 44(12), pp. 1302–1311. doi: 10.1021/ar200098t.
- Butov, D. A. *et al.* (2013) 'Randomized, placebo-controlled Phase II trial of heat-killed *Mycobacterium vaccae* (Immodulon batch) formulated as an oral pill (V7)', *Immunotherapy*, 5(10), pp. 1047–1054. doi: 10.2217/imt.13.110.
- Cahová, H. *et al.* (2015) 'NAD captureSeq indicates NAD as a bacterial cap for a subset of regulatory RNAs', *Nature*, 519(7543), pp. 374–377. doi: 10.1038/nature14020.
- Campbell, E. A. *et al.* (2001a) 'Structural mechanism for rifampicin inhibition of bacterial RNA polymerase', *Cell*. doi: 10.1016/S0092-8674(01)00286-0.
- Campbell, E. A. *et al.* (2001b) 'Structural mechanism for rifampicin inhibition of bacterial RNA polymerase', *Cell*, 104(6), pp. 901–912. doi: 10.1016/S0092-8674(01)00286-0.

- Capparelli, R. *et al.* (2007) 'Experimental phage therapy against *Staphylococcus aureus* in mice.', *Antimicrobial agents and chemotherapy*. American Society for Microbiology (ASM), 51(8), pp. 2765–73. doi: 10.1128/AAC.01513-06.
- Castello, A., Hentze, M. W. and Preiss, T. (2015) 'Metabolic Enzymes Enjoying New Partnerships as RNA-Binding Proteins', *Trends in Endocrinology and Metabolism*, pp. 746–757. doi: 10.1016/j.tem.2015.09.012.
- Catalão, M. J. *et al.* (2011) 'A Second Endolysin Gene Is Fully Embedded In-Frame with the lysA Gene of Mycobacteriophage Ms6', *PLoS ONE*. Edited by E. S. F. W. Abdelhay. Public Library of Science, 6(6), p. e20515. doi: 10.1371/journal.pone.0020515.
- Catalão, M. J. and Pimentel, M. (2018) 'Mycobacteriophage Lysis Enzymes: Targeting the Mycobacterial Cell Envelope.', *Viruses*. Multidisciplinary Digital Publishing Institute (MDPI), 10(8). doi: 10.3390/v10080428.
- Cayrol, B. *et al.* (2009) 'Auto-assembly of *E. coli* DsrA small noncoding RNA: Molecular characteristics and functional consequences', *RNA Biology*, 6(4), pp. 434–445. doi: 10.4161/rna.6.4.8949.
- Cerritelli, S. M. and Crouch, R. J. (2009) 'Ribonuclease H: the enzymes in eukaryotes', *FEBS Journal*, 276(6), pp. 1494–1505. doi: 10.1111/j.1742-4658.2009.06908.x.
- Ceska, T. A. *et al.* (1996) 'A helical arch allowing single-stranded DNA to thread through T5 5'-exonuclease', *Nature*. Nature Publishing Group, 382(6586), pp. 90–93. doi: 10.1038/382090a0.
- Chang, D. D. and Clayton, D. A. (1985) 'Priming of human mitochondrial DNA replication occurs at the light-strand promoter', *Proceedings of the National Academy of Sciences of the United States of America*, 82(2), pp. 351–355. doi: 10.1073/pnas.82.2.351.
- Chatterji, D. *et al.* (2007) 'The role of the omega subunit of RNA polymerase in expression of the *relA* gene in *Escherichia coli*', *FEMS Microbiology Letters*. John Wiley & Sons, Ltd (10.1111), 267(1), pp. 51–55. doi: 10.1111/j.1574-6968.2006.00532.x.
- Chen, Y. G. *et al.* (2009) 'LC/MS analysis of cellular RNA reveals NAD-linked RNA', *Nature Chemical Biology*. United States, 5(12), pp. 879–881. doi: 10.1038/nchembio.235.
- Cheng, Q. *et al.* (2005) 'Removal of Group B Streptococci Colonizing the Vagina and Oropharynx of Mice with a Bacteriophage Lytic Enzyme', *Antimicrobial Agents and Chemotherapy*. American Society for Microbiology Journals, 49(1), pp. 111–117. doi: 10.1128/AAC.49.1.111-117.2005.
- Chhibber, S., Kaur, S. and Kumari, S. (2008) 'Therapeutic potential of bacteriophage in treating *Klebsiella pneumoniae* B5055-mediated lobar pneumonia in mice', *Journal of Medical Microbiology*, 57(12), pp. 1508–1513. doi: 10.1099/jmm.0.2008/002873-0.
- Cho, B.-K. *et al.* (2014) 'Genome-scale reconstruction of the sigma factor network in *Escherichia coli*: topology and functional states', *BMC Biology*. BioMed Central, 12(1), p. 4. doi: 10.1186/1741-7007-12-4.
- Clark, J. R. (2015) 'Bacteriophage therapy: history and future prospects', *Future Virology*. Future Medicine Ltd London, UK, 10(4), pp. 449–461. doi: 10.2217/fvl.15.3.
- Coffey, B. *et al.* (2011) 'Assessment of *Escherichia coli* O157:H7-specific bacteriophages e11/2 and e4/1c in model broth and hide environments', *International Journal of Food*

- Microbiology*. Elsevier, 147(3), pp. 188–194. doi: 10.1016/J.IJFOODMICRO.2011.04.001.
- Colland, F. *et al.* (2002) ‘The interaction between sigmaS, the stationary phase sigma factor, and the core enzyme of *Escherichia coli* RNA polymerase’, *Genes to Cells*. John Wiley & Sons, Ltd (10.1111), 7(3), pp. 233–247. doi: 10.1046/j.1365-2443.2002.00517.x.
- Comas, I. *et al.* (2012) ‘Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes’, *Nature Genetics*. Nature Publishing Group, 44(1), pp. 106–110. doi: 10.1038/ng.1038.
- Cortes, T. *et al.* (2013) ‘Genome-wide Mapping of Transcriptional Start Sites Defines an Extensive Leaderless Transcriptome in *Mycobacterium tuberculosis*’, *Cell Reports*, 5(4), pp. 1121–1131. doi: 10.1016/j.celrep.2013.10.031.
- Cunningham, L., Gruer, M. J. and Guest, J. R. (1997) ‘Transcriptional regulation of the aconitase genes (acnA and acnB) of *Escherichia coli*’, *Microbiology*, 143(12), pp. 3795–3805. doi: 10.1099/00221287-143-12-3795.
- Cusick, M. F., Libbey, J. E. and Fujinami, R. S. (2012) ‘Molecular mimicry as a mechanism of autoimmune disease.’, *Clinical reviews in allergy & immunology*. NIH Public Access, 42(1), pp. 102–11. doi: 10.1007/s12016-011-8294-7.
- Czyz, A. *et al.* (2014) ‘Mycobacterial RNA polymerase requires a U-tract at intrinsic terminators and is aided by NusG at suboptimal terminators.’, *mBio*. American Society for Microbiology, 5(2), p. e00931. doi: 10.1128/mBio.00931-14.
- D’Herelle, F. (1929) ‘Studies Upon Asiatic Cholera.’, *The Yale journal of biology and medicine*. Yale Journal of Biology and Medicine, 1(4), pp. 195–219. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21433426> (Accessed: 17 September 2019).
- D’Heygère, F. *et al.* (2015) ‘ATP-dependent motor activity of the transcription termination factor Rho from *Mycobacterium tuberculosis*.’, *Nucleic acids research*. Oxford University Press, 43(12), pp. 6099–111. doi: 10.1093/nar/gkv505.
- D’Souza, A. R. and Minczuk, M. (2018) ‘Mitochondrial transcription and translation: overview.’, *Essays in biochemistry*. Portland Press Limited, 62(3), pp. 309–320. doi: 10.1042/EBC20170102.
- Dal Molin, M. *et al.* (2018) ‘Molecular Mechanisms of Intrinsic Streptomycin Resistance in *Mycobacterium abscessus*.’, *Antimicrobial agents and chemotherapy*. American Society for Microbiology Journals, 62(1), pp. e01427-17. doi: 10.1128/AAC.01427-17.
- Dams, D. and Briers, Y. (2019) ‘Enzybiotics: Enzyme-Based Antibacterials as Therapeutics’, in: Springer, Singapore, pp. 233–253. doi: 10.1007/978-981-13-7709-9_11.
- Dawson, R. *et al.* (2009) ‘Immunomodulation with Recombinant Interferon- γ 1b in Pulmonary Tuberculosis’, *PLoS ONE*. Edited by O. Eickelberg. Public Library of Science, 4(9), p. e6984. doi: 10.1371/journal.pone.0006984.
- Deegan, B. J. *et al.* (2011) ‘Structural and thermodynamic consequences of the replacement of zinc with environmental metals on estrogen receptor α -DNA interactions.’, *Journal of molecular recognition : JMR*. NIH Public Access, 24(6), pp. 1007–17. doi: 10.1002/jmr.1148.
- Degen, D. *et al.* (2014) ‘Transcription inhibition by the depsipeptide antibiotic salinamide A’, *eLife*, 3, p. e02451. doi: 10.7554/eLife.02451.
- Dehbi, M. *et al.* (2009) ‘Inhibition of Transcription in *Staphylococcus aureus* by a Primary

Sigma Factor-Binding Polypeptide from Phage G1 □', 191(12), pp. 3763–3771. doi: 10.1128/JB.00241-09.

Dennis, P. P. (1977) 'Transcription patterns of adjacent segments on the chromosome of *Escherichia coli* containing genes coding for four 50 S ribosomal proteins and the β and β' subunits of RNA polymerase', *Journal of Molecular Biology*. Academic Press, 115(4), pp. 603–620. doi: 10.1016/0022-2836(77)90105-X.

Deshpande, A. P. and Patel, S. S. (2014) 'Interactions of the yeast mitochondrial RNA polymerase with the +1 and +2 promoter bases dictate transcription initiation efficiency', *Nucleic Acids Research*, 42(18), pp. 11721–11732. doi: 10.1093/nar/gku868.

Deutscher, M. P. (2003) 'Degradation of Stable RNA in Bacteria', *Journal of Biological Chemistry*, pp. 45041–45044. doi: 10.1074/jbc.R300031200.

Deutscher, M. P. (2009) 'Chapter 9 Maturation and Degradation of Ribosomal RNA in Bacteria', *Progress in Molecular Biology and Translational Science*. Academic Press, 85, pp. 369–391. doi: 10.1016/S0079-6603(08)00809-X.

Dewar, J. M. and Walter, J. C. (2017) 'Mechanisms of DNA replication termination', *Nature Reviews Molecular Cell Biology*, 18(8), pp. 507–516. doi: 10.1038/nrm.2017.42.

Dimude, J. U. *et al.* (2016) 'Replication Termination: Containing Fork Fusion-Mediated Pathologies in *Escherichia coli*.', *Genes*. Multidisciplinary Digital Publishing Institute (MDPI), 7(8). doi: 10.3390/genes7080040.

Domingo-Calap, P. and Delgado-Martínez, J. (2018) 'Bacteriophages: Protagonists of a Post-Antibiotic Era.', *Antibiotics (Basel, Switzerland)*. Multidisciplinary Digital Publishing Institute (MDPI), 7(3). doi: 10.3390/antibiotics7030066.

Donnelly-Wu, M. K., Jacobs, W. R. and Hatfull, G. F. (1993) 'Superinfection immunity of mycobacteriophage L5: applications for genetic transformation of mycobacteria', *Molecular Microbiology*. John Wiley & Sons, Ltd (10.1111), 7(3), pp. 407–417. doi: 10.1111/j.1365-2958.1993.tb01132.x.

Drakulic, S., Cuellar, J. and Sousa, R. (2018) 'The Mitochondrial Transcription Machinery', in: Springer, Cham, pp. 1–15. doi: 10.1007/978-3-319-78190-7_1.

Duggin, I. G. and Bell, S. D. (2009) 'Termination Structures in the *Escherichia coli* Chromosome Replication Fork Trap', *Journal of Molecular Biology*, 387(3), pp. 532–539. doi: 10.1016/j.jmb.2009.02.027.

Dye, C. and Espinal, M. A. (2001) 'Will tuberculosis become resistant to all antibiotics?', *Proceedings. Biological sciences*. The Royal Society, 268(1462), pp. 45–52. doi: 10.1098/rspb.2000.1328.

Edwards, A. N. *et al.* (2011) 'Circuitry Linking the Csr and Stringent Response Global Regulatory Systems', *Molecular Cell*, 80(6), pp. 1561–1580. doi: 10.1111/j.1365-2958.2011.07663.x.Circuitry.

Espeli, O. *et al.* (2003) 'Temporal Regulation of Topoisomerase IV Activity in *E. coli*', *Molecular Cell*, 11(1), pp. 189–201. doi: 10.1016/S1097-2765(03)00013-3.

Fan, J. *et al.* (2016) 'Reconstruction of bacterial transcription-coupled repair at single-molecule resolution', *Nature*. Nature Publishing Group, 536(7615), pp. 234–237. doi: 10.1038/nature19080.

- Fan, X. *et al.* (2016) 'The Global Reciprocal Reprogramming between Mycobacteriophage SWU1 and Mycobacterium Reveals the Molecular Strategy of Subversion and Promotion of Phage Infection.', *Frontiers in microbiology*. Frontiers Media SA, 7, p. 41. doi: 10.3389/fmicb.2016.00041.
- Fantappiè, L. *et al.* (2009) 'The RNA chaperone Hfq is involved in stress response and virulence in *Neisseria meningitidis* and is a pleiotropic regulator of protein expression.', *Infection and immunity*. American Society for Microbiology Journals, 77(5), pp. 1842–53. doi: 10.1128/IAI.01216-08.
- Faure, G. *et al.* (2016) 'Role of mRNA structure in the control of protein folding.', *Nucleic acids research*. Oxford University Press, 44(22), pp. 10898–10911. doi: 10.1093/nar/gkw671.
- Fedoroff, O. Y., Salazar, M. and Reid, B. R. (1993) 'Structure of a DNA : RNA Hybrid Duplex: Why RNase H Does Not Cleave Pure RNA', *Journal of Molecular Biology*. Academic Press, 233(3), pp. 509–523. doi: 10.1006/JMBI.1993.1528.
- Feijó Delgado, F. *et al.* (2013) 'Intracellular water exchange for measuring the dry mass, water mass and changes in chemical composition of living cells.', *PloS one*. Public Library of Science, 8(7), p. e67590. doi: 10.1371/journal.pone.0067590.
- Foley, P. L. *et al.* (2015) 'Specificity and Evolutionary Conservation of the *Escherichia coli* RNA Pyrophosphohydrolase RppH', *Journal of Biological Chemistry*, 290(15), pp. 9478–9486. doi: 10.1074/jbc.M114.634659.
- Ford, M. E., Sarkis, G. J., *et al.* (1998) 'Genome structure of mycobacteriophage D29: implications for phage evolution', *Journal of Molecular Biology*. Academic Press, 279(1), pp. 143–164. doi: 10.1006/JMBI.1997.1610.
- Ford, M. E., Stenstrom, C., *et al.* (1998) 'Mycobacteriophage TM4: Genome structure and gene expression', *Tubercle and Lung Disease*, 79(2), pp. 63–73. doi: 10.1054/tuld.1998.0007.
- Forti, F. *et al.* (2011) 'Isolation of conditional expression mutants in *Mycobacterium tuberculosis* by transposon mutagenesis', *Tuberculosis*, 91(6), pp. 569–578. doi: 10.1016/j.tube.2011.07.004.
- Fossum, S., Crooke, E. and Skarstad, K. (2007) 'Organization of sister origins and replisomes during multifork DNA replication in *Escherichia coli*.', *The EMBO journal*. European Molecular Biology Organization, 26(21), pp. 4514–22. doi: 10.1038/sj.emboj.7601871.
- Freel, K. C., Friedrich, A. and Schacherer, J. (2015) 'Mitochondrial genome evolution in yeasts: An all-encompassing view', *FEMS Yeast Research*, 15(4), pp. 1–9. doi: 10.1093/femsyr/fov023.
- Frick, D. N. and Bessman, M. J. (1995) 'Cloning, purification, and properties of a novel NADH pyrophosphatase. Evidence for a nucleotide pyrophosphatase catalytic domain in MutT-like enzymes', *Journal of Biological Chemistry*, 270(4), pp. 1529–1534. doi: 10.1074/jbc.270.4.1529.
- Frick, D. N., Kumar, S. and Richardson, C. C. (1999) 'Interaction of Ribonucleoside Triphosphates with the Gene 4 Primase of Bacteriophage T7', *Journal of Biological Chemistry*, 274(50), pp. 35899–35907. doi: 10.1074/jbc.274.50.35899.
- Frindert, J. *et al.* (2018) 'Identification, Biosynthesis, and Decapping of NAD-Capped RNAs in *B. subtilis*', *Cell Reports*. Cell Press, 24(7), pp. 1890–1901.e8. doi: 10.1016/J.CELREP.2018.07.047.

- Fruth, M. *et al.* (2014) 'Binding mode characterization of novel RNA polymerase inhibitors using a combined biochemical and NMR approach', *ACS Chemical Biology*, 9(11), pp. 2656–2663. doi: 10.1021/cb5005433.
- Fukushima, S. *et al.* (2007) 'Reassessment of the in vivo functions of DNA polymerase I and RNase H in bacterial cell growth.', *Journal of bacteriology*. American Society for Microbiology (ASM), 189(23), pp. 8575–83. doi: 10.1128/JB.00653-07.
- Furuichi, Y. (2015) 'Discovery of m(7)G-cap in eukaryotic mRNAs.', *Proceedings of the Japan Academy. Series B, Physical and biological sciences*. The Japan Academy, 91(8), pp. 394–409. doi: 10.2183/pjab.91.394.
- Gaal, T. *et al.* (2006) 'Crl facilitates RNA polymerase holoenzyme formation.', *Journal of bacteriology*. American Society for Microbiology Journals, 188(22), pp. 7966–70. doi: 10.1128/JB.01266-06.
- García-Gómez, S. *et al.* (2013) 'PrimPol, an archaic primase/polymerase operating in human cells.', *Molecular cell*. Elsevier, 52(4), pp. 541–53. doi: 10.1016/j.molcel.2013.09.025.
- Garcia, M., Pimentel, M. and Moniz-Pereira, J. (2002) 'Expression of Mycobacteriophage Ms6 lysis genes is driven by two sigma(70)-like promoters and is dependent on a transcription termination signal present in the leader RNA.', *Journal of bacteriology*. American Society for Microbiology (ASM), 184(11), pp. 3034–43. doi: 10.1128/JB.184.11.3034-3043.2002.
- Genes - Gene Synthesis / Twist Bioscience*. Available at: <https://www.twistbioscience.com/products/genes> (Accessed: 30 October 2019).
- Gentry, D. R. and Burgess, R. R. (1989) 'rpoZ, encoding the omega subunit of *Escherichia coli* RNA polymerase, is in the same operon as spoT.', *Journal of bacteriology*. American Society for Microbiology (ASM), 171(3), pp. 1271–7. doi: 10.1128/jb.171.3.1271-1277.1989.
- Gerhard E, Wagner H, N. K. (1986) 'Structural analysis of an RNA molecule involved in replication control of plasmid R1', *Nucleic Acids Research*, 14(22), pp. 8919–8932. doi: 10.1093/nar/gkn907.
- Ghosh, G. *et al.* (2018) 'A bacteriophage capsid protein is an inhibitor of a conserved transcription terminator of various bacterial pathogens', *Journal of Bacteriology*, 200(1), pp. 1–16. doi: 10.1128/JB.00380-17.
- Gil, F. *et al.* (2008) 'The lytic cassette of mycobacteriophage Ms6 encodes an enzyme with lipolytic activity', *Microbiology*. Microbiology Society, 154(5), pp. 1364–1371. doi: 10.1099/mic.0.2007/014621-0.
- Gloor, J. W., Balakrishnan, L. and Bambara, R. A. (2010) 'Flap endonuclease 1 mechanism analysis indicates flap base binding prior to threading.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 285(45), pp. 34922–31. doi: 10.1074/jbc.M110.165902.
- Gnatt, A. L. *et al.* (2001) 'Structural Basis of Transcription: An RNA Polymerase II Elongation Complex at 3.3 Å Resolution', *Science*, 292(5523), pp. 1876–1882. doi: 10.1126/science.1059495.
- Godson, G. N. *et al.* (2000) 'Identification of the Magnesium Ion Binding Site in the Catalytic Center of *Escherichia coli* Primase by Iron Cleavage [†]', *Biochemistry*, 39(2), pp. 332–339. doi: 10.1021/bi9916628.

- Gorodetsky, A. A., Boal, A. K. and Barton, J. K. (2006) 'Direct electrochemistry of endonuclease III in the presence and absence of DNA', *Journal of the American Chemical Society*. American Chemical Society, 128(37), pp. 12082–12083. doi: 10.1021/ja064784d.
- Gottesman, M. E. and Weisberg, R. A. (1995) 'Termination and antitermination of transcription in temperate bacteriophages', *Seminars in Virology*. Academic Press, 6(1), pp. 35–42. doi: 10.1016/S1044-5773(05)80007-1.
- Grudzien-Nogalska, E. *et al.* (2018) "'NAD-capQ" detection and quantitation of NAD caps', *RNA*. Cold Spring Harbor Laboratory Press, 24(10), pp. 1418–1425. doi: 10.1261/rna.067686.118.
- Grudzien-Nogalska, E. and Kiledjian, M. (2017) 'New insights into decapping enzymes and selective mRNA decay', *Wiley Interdisciplinary Reviews: RNA*, pp. 1–18. doi: 10.1002/wrna.1379.
- Guberman, J. M. *et al.* (2008) 'PSICIC: Noise and Asymmetry in Bacterial Division Revealed by Computational Image Analysis at Sub-Pixel Resolution', *PLoS Computational Biology*. Edited by M. Feingold, 4(11), p. e1000233. doi: 10.1371/journal.pcbi.1000233.
- Guild, N. *et al.* (1988) 'Transcriptional activation of bacteriophage T4 middle promoters by the MotA protein', *Journal of Molecular Biology*. Academic Press, 199(2), pp. 241–258. doi: 10.1016/0022-2836(88)90311-7.
- Gusarov, I. and Nudler, E. (1999) 'The Mechanism of Intrinsic Transcription Termination', *Molecular Cell*. Cell Press, 3(4), pp. 495–504. doi: 10.1016/S1097-2765(00)80477-3.
- Hamdan, S. M. *et al.* (2007) 'Dynamic DNA Helicase-DNA Polymerase Interactions Assure Processive Replication Fork Movement', *Molecular Cell*. Cell Press, 27(4), pp. 539–549. doi: 10.1016/J.MOLCEL.2007.06.020.
- Han, Q. and Eiteman, M. A. (2018) 'Enhancement of NAD(H) pool for formation of oxidized biochemicals in *Escherichia coli*', *Journal of Industrial Microbiology & Biotechnology*. Springer International Publishing, 45(11), pp. 939–950. doi: 10.1007/s10295-018-2072-y.
- Harrington, J. J. and Lieber, M. R. (1994) 'The characterization of a mammalian DNA structure-specific endonuclease.', *The EMBO journal*, 13(5), pp. 1235–46. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8131753> (Accessed: 2 October 2019).
- Hatfull, Graham, Russell, Dan, Jacobs-Sera, Debbie, Pope, Welkin, Sivanathan, Viknesh, Tse, E. (2016) *The Actinobacteriophage Database / Home, HHMI at PhagesDB.org*. Available at: <https://phagesdb.org/> (Accessed: 30 October 2019).
- Hatfull, G. F. (2008) 'Bacteriophage genomics', *Current Opinion in Microbiology*. doi: 10.1016/j.mib.2008.09.004.
- Hatfull, G. F. (2010a) 'Mycobacteriophages: Genes and Genomes', *Annual Review of Microbiology*. doi: 10.1146/annurev.micro.112408.134233.
- Hatfull, G. F. (2010b) 'Mycobacteriophages: Genes and Genomes', *Annual Review of Microbiology*. Annual Reviews, 64(1), pp. 331–356. doi: 10.1146/annurev.micro.112408.134233.
- Hatfull, G. F. (2014a) 'Molecular Genetics of Mycobacteriophages.', *Microbiology spectrum*. NIH Public Access, 2(2), pp. 1–36. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25328854> (Accessed: 3 October 2019).

- Hatfull, G. F. (2014b) 'Mycobacteriophages: windows into tuberculosis.', *PLoS pathogens*. Public Library of Science, 10(3), p. e1003953. doi: 10.1371/journal.ppat.1003953.
- Hatfull, G. F. (2018) 'Mycobacteriophages', *Microbiology Spectrum*, 6(5). doi: 10.1128/microbiolspec.GPP3-0026-2018.
- Hatfull and Jacobs (eds) (2014) *Molecular Genetics of Mycobacteria, Second Edition*. American Society of Microbiology. doi: 10.1128/9781555818845.
- Hatfull, G. F. and Sarkis, G. J. (1993) 'DNA sequence, structure and gene expression of mycobacteriophage L5: a phage system for mycobacterial genetics', *Molecular Microbiology*. John Wiley & Sons, Ltd (10.1111), 7(3), pp. 395–405. doi: 10.1111/j.1365-2958.1993.tb01131.x.
- Herbert, K. M. *et al.* (2010) '*E. coli* NusG inhibits backtracking and accelerates pause-free transcription by promoting forward translocation of RNA polymerase', *Journal of Molecular Biology*. doi: 10.1016/j.jmb.2010.03.051.
- Hermoso, J. A., García, J. L. and García, P. (2007) 'Taking aim on bacterial pathogens: from phage therapy to enzybiotics', *Current Opinion in Microbiology*, pp. 461–472. doi: 10.1016/j.mib.2007.08.002.
- Hiasa, H. and Marians, K. J. (1994) 'Tus prevents overreplication of oriC plasmid DNA.', *The Journal of biological chemistry*, 269(43), pp. 26959–68. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7929435> (Accessed: 2 October 2019).
- Hiasa, H. and Marians, K. J. (1996) 'Two Distinct Modes of Strand Unlinking during θ -Type DNA Replication', *Journal of Biological Chemistry*, 271(35), pp. 21529–21535. doi: 10.1074/jbc.271.35.21529.
- Hillen, H. S., Temiakov, D. and Cramer, P. (2018) 'Structural basis of mitochondrial transcription', *Nature Structural & Molecular Biology* 25:9. Nature Publishing Group, 25(9), pp. 754–765. doi: 10.1038/s41594-018-0122-9.
- Hinton, D. M. *et al.* (1996) 'Bacteriophage T4 middle transcription system: T4-modified RNA polymerase; AsiA, a σ 70 binding protein; and transcriptional activator MotA', *Methods in Enzymology*. Academic Press, 274, pp. 43–57. doi: 10.1016/S0076-6879(96)74007-7.
- Höfer, K. *et al.* (2016) 'Structure and function of the bacterial decapping enzyme NudC', *Nature Chemical Biology*, 12(9), pp. 730–734. doi: 10.1038/nchembio.2132.
- Hoffmann, C. *et al.* (2008) 'Disclosure of the mycobacterial outer membrane: cryo-electron tomography and vitreous sections reveal the lipid bilayer structure.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 105(10), pp. 3963–7. doi: 10.1073/pnas.0709530105.
- Hofman, S. *et al.* (2016) 'Emerging drugs and alternative possibilities in the treatment of tuberculosis', *Expert Opinion on Emerging Drugs*. Taylor & Francis, 21(1), pp. 103–116. doi: 10.1517/14728214.2016.1151000.
- Hoppensteadt, F. C. (1989) 'Synchronization of Bacterial Culture Growth', in: Springer, Berlin, Heidelberg, pp. 16–22. doi: 10.1007/978-3-642-46693-9_2.
- Hosfield, D. J. *et al.* (1998) 'Structure of the DNA repair and replication endonuclease and exonuclease FEN-1: coupling DNA and PCNA binding to FEN-1 activity.', *Cell*, 95(1), pp. 135–46. doi: 10.1016/s0092-8674(00)81789-4.

- Hou, C., Biswas, T. and Tsodikov, O. V. (2018) 'Structures of the Catalytic Domain of Bacterial Primase DnaG in Complexes with DNA Provide Insight into Key Priming Events', *Biochemistry*. American Chemical Society, 57(14), pp. 2084–2093. doi: 10.1021/acs.biochem.8b00036.
- Hu, Y. *et al.* (2012) 'Mycobacterium tuberculosis RbpA protein is a new type of transcriptional activator that stabilizes the σ A -containing RNA polymerase holoenzyme', *Nucleic Acids Research*, 40(14), pp. 6547–6557. doi: 10.1093/nar/gks346.
- Huang, F. (2003) 'Efficient incorporation of CoA, NAD and FAD into RNA by in vitro transcription', *Nucleic Acids Research*, 31(3), pp. 8e – 8. doi: 10.1093/nar/gng008.
- Huang, F., Bugg, C. W. and Yarus, M. (2000) 'RNA-catalyzed CoA, NAD, and FAD synthesis from phosphopantetheine, NMN, and FMN', *Biochemistry*, 39(50), pp. 15548–15555. doi: 10.1021/bi002061f.
- Hubin, E. A. *et al.* (2015) 'Structural, functional, and genetic analyses of the actinobacterial transcription factor RbpA', *Proceedings of the National Academy of Sciences*, 112(23), pp. 7171–7176. doi: 10.1073/pnas.1504942112.
- Hubin, Elizabeth A. *et al.* (2017) 'Structural insights into the mycobacteria transcription initiation complex from analysis of X-ray crystal structures', *Nature Communications*. Nature Publishing Group, 8(1), p. 16072. doi: 10.1038/ncomms16072.
- Hubin, Elizabeth A *et al.* (2017) 'Structure and function of the mycobacterial transcription initiation complex with the essential regulator RbpA', *eLife*, 6. doi: 10.7554/eLife.22520.
- Ilic, S. *et al.* (2016) 'Identification of DNA primase inhibitors via a combined fragment-based and virtual screening', *Scientific Reports*, 6(36322), pp. 1–10. doi: 10.1038/srep36322.
- Ilic, S. *et al.* (2018) 'DnaG Primase—A Target for the Development of Novel Antibacterial Agents', *Antibiotics*. Multidisciplinary Digital Publishing Institute, 7(3), p. 72. doi: 10.3390/antibiotics7030072.
- Imashimizu, M. *et al.* (2014) 'Transcription elongation. Heterogeneous tracking of RNA polymerase and its biological implications.', *Transcription*. Taylor & Francis, 5(1), p. e28285. doi: 10.4161/trns.28285.
- Ito, N. *et al.* (2003) 'Crystal structure of the *Pyrococcus horikoshii* DNA primase-UTP complex: implications for the mechanism of primer synthesis.', *Genes to cells: devoted to molecular & cellular mechanisms*, 8(12), pp. 913–23. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/14750947> (Accessed: 1 October 2019).
- Jacob, F., Brenner, S. and Cuzin, F. (1963) 'On the Regulation of DNA Replication in Bacteria', *Cold Spring Harbor Symposia on Quantitative Biology*. Cold Spring Harbor Laboratory Press, 28(0), pp. 329–348. doi: 10.1101/SQB.1963.028.01.048.
- Jacob, S. T. (1995) 'Regulation of ribosomal gene transcription.', *The Biochemical journal*. Portland Press Ltd, 306 (Pt 3)(Pt 3), pp. 617–26. doi: 10.1042/bj3060617.
- Jacobs-Sera, D. *et al.* (2012) 'On the nature of mycobacteriophage diversity and host preference.', *Virology*. NIH Public Access, 434(2), pp. 187–201. doi: 10.1016/j.virol.2012.09.026.
- Jaschke, A. *et al.* (2016) 'Cap-like structures in bacterial RNA and epitranscriptomic modification.', *Current opinion in microbiology*. England, 30, pp. 44–49. doi:

10.1016/j.mib.2015.12.009.

Jäschke, A. *et al.* (2016) 'Cap-like structures in bacterial RNA and epitranscriptomic modification', *Current Opinion in Microbiology*, 30, pp. 44–49. doi: 10.1016/j.mib.2015.12.009.

Jeffares, D. C., Poole, A. M. and Penny, D. (1998) 'Relics from the RNA world', *Journal of molecular evolution*, 46(1), pp. 18–36. doi: 10.1007/PL00006275.

Jiao, X. *et al.* (2017) '5' End Nicotinamide Adenine Dinucleotide Cap in Human Cells Promotes RNA Decay through DXO-Mediated deNADding', *Cell*, 168(6), pp. 1015–1027.e10. doi: 10.1016/j.cell.2017.02.019.

Jo, A., Ding, T. and Ahn, J. (2016) 'Synergistic antimicrobial activity of bacteriophages and antibiotics against *Staphylococcus aureus*.', *Food science and biotechnology*. Springer, 25(3), pp. 935–940. doi: 10.1007/s10068-016-0153-0.

Johnson, S. K., Bhattacharyya, S. and Griep, M. A. (2000) 'DnaB helicase stimulates primer synthesis activity on short oligonucleotide templates', *Biochemistry*. American Chemical Society, 39(4), pp. 736–744. doi: 10.1021/bi991554l.

Jones, W. and White, A. (1968) 'Lysogeny in mycobacteria. I. Conversion of colony morphology, nitrate reductase activity, and Tween 80 hydrolysis of *Mycobacterium* sp. ATCC 607 associated with lysogeny', *Canadian Journal of Microbiology*, 14(5), pp. 551–555. doi: 10.1139/m68-093.

Jørgensen, M. G. *et al.* (2013) 'Dual function of the McaS small RNA in controlling biofilm formation Dual function of the McaS small RNA in controlling biofilm formation', *Genes & Development*, 27, pp. 1132–1145. doi: 10.1101/gad.214734.113.

Julius, C., Riaz-Bradley, A. and Yuzenkova, Y. (2018) 'RNA capping by mitochondrial and multi-subunit RNA polymerases', *Transcription*. Taylor & Francis, 1264, pp. 1–13. doi: 10.1080/21541264.2018.1456258.

Julius, C. and Yuzenkova, Y. (2017) 'Bacterial RNA polymerase caps RNA with various cofactors and cell wall precursors', *Nucleic acids research*. Oxford University Press, 45(14), pp. 8282–8290. doi: 10.1093/nar/gkx452.

Julius, C. and Yuzenkova, Y. (2019) 'Noncanonical RNA-capping: Discovery, mechanism, and physiological role debate', *Wiley Interdisciplinary Reviews: RNA*. John Wiley & Sons, Ltd, 10(2), p. e1512. doi: 10.1002/wrna.1512.

Kalate, R. N., Kulkarni, B. D. and Nagaraja, V. (2002) 'Analysis of DNA curvature distribution in mycobacterial promoters using theoretical models', *Biophysical Chemistry*. doi: 10.1016/S0301-4622(02)00124-2.

Kang, Y.-H., Lee, C.-H. and Seo, Y.-S. (2010) 'Dna2 on the road to Okazaki fragment processing and genome stability in eukaryotes', *Critical Reviews in Biochemistry and Molecular Biology*, 45(2), pp. 71–96. doi: 10.3109/10409230903578593.

Kazlauskas, D. *et al.* (2018) 'Novel Families of Archaeo-Eukaryotic Primases Associated with Mobile Genetic Elements of Bacteria and Archaea', *Journal of Molecular Biology*, 430(5), pp. 737–750. doi: 10.1016/j.jmb.2017.11.014.

Keck, J. L. *et al.* (2000) 'Structure of the RNA polymerase domain of *E. coli* primase.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 287(5462),

pp. 2482–6. doi: 10.1126/science.287.5462.2482.

Kelch, B. A. *et al.* (2011) ‘How a DNA polymerase clamp loader opens a sliding clamp.’, *Science (New York, N.Y.)*. NIH Public Access, 334(6063), pp. 1675–80. doi: 10.1126/science.1211884.

Kelch, B. A. *et al.* (2012) ‘Clamp loader ATPases and the evolution of DNA replication machinery’, *BMC Biology*. BioMed Central, 10(1), p. 34. doi: 10.1186/1741-7007-10-34.

Keller, K. E., Cavanaugh, N. and Kuchta, R. D. (2008) ‘Interaction of herpes primase with the sugar of a NTP.’, *Biochemistry*. NIH Public Access, 47(34), pp. 8977–84. doi: 10.1021/bi8008467.

Kelly, R. B. *et al.* (1969) ‘Excision of thymine dimers and other mismatched sequences by DNA polymerase of *Escherichia coli*’, *Nature*, 224(5218), pp. 495–501. doi: 10.1038/224495a0.

Kiledjian, M. (2018) ‘Eukaryotic RNA 5'-End NAD⁺Capping and DeNADding’, *Trends in Cell Biology*, pp. 454–464. doi: 10.1016/j.tcb.2018.02.005.

Kilkenny, M. L. *et al.* (2013) ‘Structures of human primase reveal design of nucleotide elongation site and mode of Pol α tethering.’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 110(40), pp. 15961–6. doi: 10.1073/pnas.1311185110.

Kim, M. *et al.* (2018) ‘Phage-Antibiotic Synergy via Delayed Lysis’, *Appl. Environ. Microbiol.* American Society for Microbiology, 84(22), pp. e02085-18. doi: 10.1128/AEM.02085-18.

Kirk, B. W. *et al.* (1997) ‘Eucaryotic DNA Primase Does Not Prefer To Synthesize Primers at Pyrimidine Rich DNA Sequences When Nucleoside Triphosphates Are Present at Concentrations Found in Whole Cells [†]’, *Biochemistry*, 36(22), pp. 6725–6731. doi: 10.1021/bi962630c.

Kirtania, P. *et al.* (2016) ‘Vitamin C induced DevR-dependent synchronization of *Mycobacterium smegmatis* growth and its effect on the proliferation of mycobacteriophage D29’, *FEMS Microbiology Letters*, pp. 1–9. doi: 10.1093/femsle/fnw097.

Kitani, T. *et al.* (1985) ‘Evidence that discontinuous DNA replication in *Escherichia coli* is primed by approximately 10 to 12 residues of RNA starting with a purine’, *Journal of Molecular Biology*, 184(1), pp. 45–52. doi: 10.1016/0022-2836(85)90042-7.

Klann, A. G. *et al.* (1998) ‘Characterization of the dnaG locus in *Mycobacterium smegmatis* reveals linkage of DNA replication and cell division.’, *Journal of bacteriology*. American Society for Microbiology (ASM), 180(1), pp. 65–72. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9422594> (Accessed: 25 October 2019).

Klimuk, E. *et al.* (2013) ‘Host RNA polymerase inhibitors encoded by ϕ KMV-like phages of *pseudomonas*’, *Virology*. Academic Press, 436(1), pp. 67–74. doi: 10.1016/J.VIROL.2012.10.021.

Klinge, S. *et al.* (2007) ‘An iron-sulfur domain of the eukaryotic primase is essential for RNA primer synthesis’, *Nature Structural & Molecular Biology*. Nature Publishing Group, 14(9), pp. 875–877. doi: 10.1038/nsmb1288.

Kluska, K., Adamczyk, J. and Krężel, A. (2018) ‘Metal binding properties, stability and

- reactivity of zinc fingers', *Coordination Chemistry Reviews*. Elsevier, 367, pp. 18–64. doi: 10.1016/J.CCR.2018.04.009.
- Koch, A., Mizrahi, V. and Warner, D. F. (2014) 'The impact of drug resistance on *Mycobacterium tuberculosis* physiology: what can we learn from rifampicin?', *Emerging microbes & infections*. Nature Publishing Group, 3(3), p. e17. doi: 10.1038/emi.2014.17.
- Kogoma, T. (1986) 'RNase H-defective mutants of *Escherichia coli*.', *Journal of bacteriology*. American Society for Microbiology (ASM), 166(2), pp. 361–3. doi: 10.1128/jb.166.2.361-363.1986.
- Konopka, J. B. (2012) 'N-Acetylglucosamine Functions in Cell Signaling', *Hindawai Publishing Corporation Scientifica*. Hindawi Publishing Corporation Scientifica, (Article ID 489208), pp. 1–15. doi: 10.6064/2012/489208.
- Konrad, E. B. and Lehman, I. R. (1974) 'A conditional lethal mutant of *Escherichia coli* K12 defective in the 5' leads to 3' exonuclease associated with DNA polymerase I.', *Proceedings of the National Academy of Sciences of the United States of America*, 71(5), pp. 2048–51. doi: 10.1073/pnas.71.5.2048.
- Korzheva, N. *et al.* (2000) 'A Structural Model of Transcription Elongation', *Science*, 289(5479), pp. 619–625. doi: 10.1126/science.289.5479.619.
- Kowtoniuk, W. E. *et al.* (2009) 'A chemical screen for biological small molecule-RNA conjugates reveals CoA-linked RNA.', *Proceedings of the National Academy of Sciences of the United States of America*, 106(19), pp. 7768–73. doi: 10.1073/pnas.0900528106.
- Krut, O. and Bekeredjian-Ding, I. (2018) 'Contribution of the Immune Response to Phage Therapy.', *Journal of immunology (Baltimore, Md. : 1950)*. American Association of Immunologists, 200(9), pp. 3037–3044. doi: 10.4049/jimmunol.1701745.
- Kuchta, R. D. and Stengel, G. (2010) 'Mechanism and evolution of DNA primases', *Biochimica biophysica acta*, 1804(5), pp. 1180–1189. doi: 10.1016/j.bbapap.2009.06.011.Mechanism.
- Kuchta, R. D. and Willhelm, L. (1991) 'Inhibition of DNA primase by 9-.beta.-D-arabinofuranosyladenosine triphosphate', *Biochemistry*, 30(3), pp. 797–803. doi: 10.1021/bi00217a033.
- Kulbachinskiy, A. and Mustaev, A. (2006) 'Region 3.2 of the σ subunit contributes to the binding of the 3'-initiating nucleotide in the RNA polymerase active center and facilitates promoter clearance during initiation', *Journal of Biological Chemistry*, 281(27), pp. 18273–18276. doi: 10.1074/jbc.C600060200.
- Kulka, K., Hatfull, G. and Ojha, A. K. (2012) 'Growth of *Mycobacterium tuberculosis* biofilms.', *Journal of visualized experiments : JoVE*. MyJoVE Corporation, (60). doi: 10.3791/3820.
- Kusakabe, T. and Richardson, C. C. (1997) 'Template recognition and ribonucleotide specificity of the DNA primase of bacteriophage T7.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 272(9), pp. 5943–51. doi: 10.1074/jbc.272.9.5943.
- Kuznedelov, K. *et al.* (2002) 'Structure-based analysis of RNA polymerase function: the largest subunit's rudder contributes critically to elongation complex stability and is not involved in the maintenance of RNA-DNA hybrid length', *The EMBO Journal*, 21(6), pp.

1369–1378. doi: 10.1093/emboj/21.6.1369.

Lai, J.-L. *et al.* (2018) ‘The RNA chaperone Hfq is important for the virulence, motility and stress tolerance in the phytopathogen *Xanthomonas campestris*’, *Environmental Microbiology Reports*, 10(5), pp. 542–554. doi: 10.1111/1758-2229.12657.

Lai, M.-J. *et al.* (2015) ‘Antimycobacterial Activities of Endolysins Derived From a Mycobacteriophage, BTCU-1’, *Molecules*. Multidisciplinary Digital Publishing Institute, 20(10), pp. 19277–19290. doi: 10.3390/molecules201019277.

Lambert, L. J. *et al.* (2004) ‘T4 AsiA blocks DNA recognition by remodeling sigma70 region 4.’, *The EMBO journal*. European Molecular Biology Organization, 23(15), pp. 2952–62. doi: 10.1038/sj.emboj.7600312.

Lao-Sirieix, S.-H. *et al.* (2005) ‘Structure of the heterodimeric core primase’, *Nature Structural & Molecular Biology*, 12(12), pp. 1137–1144. doi: 10.1038/nsmb1013.

Laptenko, O. *et al.* (2003) ‘Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase.’, *The EMBO journal*. European Molecular Biology Organization, 22(23), pp. 6322–34. doi: 10.1093/emboj/cdg610.

Larson, M. H. *et al.* (2008) ‘Applied Force Reveals Mechanistic and Energetic Details of Transcription Termination’, *Cell*. Cell Press, 132(6), pp. 971–982. doi: 10.1016/J.CELL.2008.01.027.

Le, T. T. *et al.* (2018) ‘Erratum: Mfd Dynamically Regulates Transcription via a Release and Catch-Up Mechanism (Cell (2018) 172(1-2) (344–357.e15)(S0092867417313648)(10.1016/j.cell.2017.11.017))’, *Cell*. Cell Press, p. 1823. doi: 10.1016/j.cell.2018.06.002.

Lee, J. and Borukhov, S. (2016) ‘Bacterial RNA Polymerase-DNA Interaction—The Driving Force of Gene Expression and the Target for Drug Action’, *Frontiers in Molecular Biosciences*. Frontiers, 3, p. 73. doi: 10.3389/fmolb.2016.00073.

Lee, S.-J. *et al.* (2012) ‘Zinc-binding Domain of the Bacteriophage T7 DNA Primase Modulates Binding to the DNA Template’, *Journal of Biological Chemistry*, 287(46), pp. 39030–39040. doi: 10.1074/jbc.M112.414151.

Lee, S.-J. and Richardson, C. C. (2005) ‘Acidic residues in the nucleotide-binding site of the bacteriophage T7 DNA primase.’, *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 280(29), pp. 26984–91. doi: 10.1074/jbc.M504817200.

Lee, S. J. and Richardson, C. C. (2001) ‘Essential lysine residues in the RNA polymerase domain of the gene 4 primase-helicase of bacteriophage T7.’, *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 276(52), pp. 49419–26. doi: 10.1074/jbc.M108443200.

Lei, Y. *et al.* (2013) ‘Functional analysis of the protein Veg, which stimulates biofilm formation in *Bacillus subtilis*.’, *Journal of bacteriology*. American Society for Microbiology Journals, 195(8), pp. 1697–705. doi: 10.1128/JB.02201-12.

Leppek, K. and Barna, M. (2019) ‘An rRNA variant to deal with stress’, *Nature Microbiology*. Nature Publishing Group, 4(3), pp. 382–383. doi: 10.1038/s41564-019-0396-7.

Leung, C. Y. (Joey) and Weitz, J. S. (2017) ‘Modeling the synergistic elimination of bacteria by phage and the innate immune system’, *Journal of Theoretical Biology*, 429, pp. 241–252.

doi: 10.1016/j.jtbi.2017.06.037.

Levine, C., Hiasa, H. and Marians, K. J. (1998) 'DNA gyrase and topoisomerase IV: biochemical activities, physiological roles during chromosome replication, and drug sensitivities.', *Biochimica et biophysica acta*, 1400(1–3), pp. 29–43. doi: 10.1016/s0167-4781(98)00126-2.

Lewis, J. A. and Hatfull, G. F. (2001) 'Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins', *Nucleic Acids Research*, 29(11), pp. 2205–2216. doi: 10.1093/nar/29.11.2205.

Lewis, J. S., Jergic, S. and Dixon, N. E. (2016) *The E. coli DNA Replication Fork, DNA Replication Across Taxa*. doi: 10.1016/bs.enz.2016.04.001.

Lieber, M. R. (1997) 'The FEN-1 family of structure-specific nucleases in eukaryotic dna replication, recombination and repair', *BioEssays*, 19(3), pp. 233–240. doi: 10.1002/bies.950190309.

Liew, L. P. *et al.* (2016) 'Hydroxyurea-Mediated Cytotoxicity Without Inhibition of Ribonucleotide Reductase', *Cell Reports*. Cell Press, 17(6), pp. 1657–1670. doi: 10.1016/J.CELREP.2016.10.024.

Linderoth, N. A. and Calendar, R. L. (1991) 'The Psu protein of bacteriophage P4 is an antitermination factor for rho-dependent transcription termination.', *Journal of bacteriology*. American Society for Microbiology (ASM), 173(21), pp. 6722–31. doi: 10.1128/jb.173.21.6722-6731.1991.

Lipps, G. *et al.* (2004) 'Structure of a bifunctional DNA primase-polymerase', *Nature Structural & Molecular Biology*, 11(2), pp. 157–162. doi: 10.1038/nsmb723.

Liu, B. *et al.* (2015) 'A primase subunit essential for efficient primer synthesis by an archaeal eukaryotic-type primase', *Nature Communications*. Nature Publishing Group, 6, pp. 1–11. doi: 10.1038/ncomms8300.

Lodish, H. *et al.* (2004) *Molecular cell biology*, W. H. Freeman. W.H. Freeman. Available at: <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Molecular+Cell+Biology#0> (Accessed: 22 August 2019).

Lopez Sanchez, M. I. G. *et al.* (2011) 'RNA processing in human mitochondria', *Cell Cycle*. Taylor & Francis, 10(17), pp. 2904–2916. doi: 10.4161/cc.10.17.17060.

Lovett, S. T. (2007) 'Polymerase Switching in DNA Replication', *Molecular Cell*. Cell Press, 27(4), pp. 523–526. doi: 10.1016/J.MOLCEL.2007.08.003.

Lu, Y. B. *et al.* (1996) 'Direct physical interaction between DnaG primase and DnaB helicase of *Escherichia coli* is necessary for optimal synthesis of primer RNA.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 93(23), pp. 12902–7. doi: 10.1073/pnas.93.23.12902.

Lundquist, R. C. and Olivera, B. M. (1982) 'Transient generation of displaced single-stranded DNA during nick translation', *Cell*. Cell Press, 31(1), pp. 53–60. doi: 10.1016/0092-8674(82)90404-4.

Lyamichev, V., Brow, M. A. D. and Dahlberg, J. E. (1993) 'Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases', *Science*, 260(5109), pp. 778–783. doi: 10.1126/science.7683443.

- Ma, C., Yang, X. and Lewis, P. J. (2016) 'Bacterial Transcription as a Target for Antibacterial Drug Development.', *Microbiology and molecular biology reviews : MMBR*. American Society for Microbiology, 80(1), pp. 139–60. doi: 10.1128/MMBR.00055-15.
- Maciąg-Dorszyńska, M., Szalewska-Pałasz, A. and Węgrzyn, G. (2013) 'Different effects of ppGpp on *Escherichia coli* DNA replication *in vivo* and *in vitro*.' , *FEBS open bio*. Wiley-Blackwell, 3, pp. 161–4. doi: 10.1016/j.fob.2013.03.001.
- Maciąg, M. *et al.* (2010) 'ppGpp inhibits the activity of *Escherichia coli* DnaG primase', *Plasmid*. Academic Press, 63(1), pp. 61–67. doi: 10.1016/J.PLASMID.2009.11.002.
- Magnusson, L. U., Farewell, A. and Nyström, T. (2005) 'ppGpp: a global regulator in *Escherichia coli*.' , *Trends in microbiology*. Elsevier, 13(5), pp. 236–42. doi: 10.1016/j.tim.2005.03.008.
- Mahapatra, S. *et al.* (2013) 'Mycobacteriophage Ms6 LysA: a peptidoglycan amidase and a useful analytical tool.' , *Applied and environmental microbiology*. American Society for Microbiology, 79(3), pp. 768–73. doi: 10.1128/AEM.02263-12.
- Malygin, A. G. and Shemyakin, M. F. (1979) 'Adenosine, NAD and FAD can initiate template-dependent RNA a synthesis catalyzed by *Escherichia Coli* RNA polymerase', *FEBS Letters*, 102(1), pp. 51–54. doi: 10.1016/0014-5793(79)80926-6.
- Mandin, P. and Gottesman, S. (2009) 'Regulating the regulator: An RNA decoy acts as an OFF switch for the regulation of an sRNA', *Genes and Development*, 23(17), pp. 1981–1985. doi: 10.1101/gad.1846609.
- Mankiewicz, E. and Béland, J. E. (1972) 'Failure to Produce Experimental Sarcoidosis in Guinea Pigs with *Mycobacterium tuberculosis* and Mycobacteriophage DS6A', *American Review of Respiratory Disease*. American Lung Association , 106(2), pp. 284–285. doi: 10.1164/arrd.1972.106.2.284.
- Manoharadas, S., Witte, A. and Bläsi, U. (2009) 'Antimicrobial activity of a chimeric enzybiotic towards *Staphylococcus aureus*', *Journal of Biotechnology*. Elsevier, 139(1), pp. 118–123. doi: 10.1016/J.JBIOTEC.2008.09.003.
- Marceau, A. H. *et al.* (2011) 'Structure of the SSB-DNA polymerase III interface and its role in DNA replication', *The EMBO Journal*, 30(20), pp. 4236–4247. doi: 10.1038/emboj.2011.305.
- Markovitz, A. (2005) 'A new *in vivo* termination function for DNA polymerase I of *Escherichia coli* K12', *Molecular Microbiology*, 55(6), pp. 1867–1882. doi: 10.1111/j.1365-2958.2005.04513.x.
- Marr, M. T. and Roberts, J. W. (2000) 'Function of Transcription Cleavage Factors GreA and GreB at a Regulatory Pause Site', *Molecular Cell*. Cell Press, 6(6), pp. 1275–1285. doi: 10.1016/S1097-2765(00)00126-X.
- Mathew, R. and Chatterji, D. (2006) 'The evolving story of the omega subunit of bacterial RNA polymerase', *Trends in Microbiology*. doi: 10.1016/j.tim.2006.08.002.
- Mauer, J. *et al.* (2017) 'Reversible methylation of m6Amin the 5' cap controls mRNA stability', *Nature*, 541(7637), pp. 371–375. doi: 10.1038/nature21022.
- McAllister, W. T. and Barrett, C. L. (1977) 'Roles of the early genes of bacteriophage T7 in shutoff of host macromolecular synthesis.' , *Journal of virology*. American Society for

- Microbiology (ASM), 23(3), pp. 543–53. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/330878> (Accessed: 3 October 2019).
- McClure, W. R. and Cech, C. L. (1978) ‘On the mechanism of rifampicin inhibition of RNA synthesis.’, *The Journal of biological chemistry*, 253(24), pp. 8949–56. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/363713> (Accessed: 5 November 2018).
- McLennan, A. G. (2006) ‘The Nudix hydrolase superfamily’, *Cellular and Molecular Life Sciences*. Birkhäuser-Verlag, 63(2), pp. 123–143. doi: 10.1007/s00018-005-5386-7.
- Meek, D. W. and Hayward, R. S. (1984) ‘Nucleotide sequence of the *rpoA-rplQ* DNA of *Escherichia coli*: a second regulatory binding site for protein S4?’, *Nucleic Acids Research*, 12(14), pp. 5813–5821. doi: 10.1093/nar/12.14.5813.
- Meek, D. W. and Hayward, R. S. (1986) ‘Direct evidence for autogenous regulation of the *Escherichia coli* genes *rpoBC* in vivo’, *MGG Molecular & General Genetics*. Springer-Verlag, 202(3), pp. 500–508. doi: 10.1007/BF00333284.
- Mengin-Lecreux, D., Flouret, B. and van Heijenoort, J. (1983) ‘Pool levels of UDP N-acetylglucosamine and UDP N-acetylglucosamine-enolpyruvate in *Escherichia coli* and correlation with peptidoglycan synthesis.’, *Journal of bacteriology*, 154(3), pp. 1284–90. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/6222035> (Accessed: 11 October 2019).
- Meysman, P. *et al.* (2014) ‘Structural properties of prokaryotic promoter regions correlate with functional features’, *PLoS ONE*. doi: 10.1371/journal.pone.0088717.
- Mishanina, T. V *et al.* (2017) ‘Trigger loop of RNA polymerase is a positional, not acid-base, catalyst for both transcription and proofreading.’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 114(26), pp. E5103–E5112. doi: 10.1073/pnas.1702383114.
- Mitchell, S. F. and Parker, R. (2015) ‘Modifications on Translation Initiation’, *Cell*. Cell Press, 163(4), pp. 796–798. doi: 10.1016/J.CELL.2015.10.056.
- Mohanty, B. K. and Kushner, S. R. (2006) ‘The majority of *Escherichia coli* mRNAs undergo post-transcriptional modification in exponentially growing cells’, *Nucleic Acids Research*, 34(19), pp. 5695–5704. doi: 10.1093/nar/gkl684.
- Mohanty, B. K. and Kushner, S. R. (2016) ‘Regulation of mRNA Decay in Bacteria’, *Annual Review of Microbiology*, 70(1), pp. 25–44. doi: 10.1146/annurev-micro-091014-104515.
- Moll, I. *et al.* (2002) ‘Leaderless mRNAs in bacteria: Surprises in ribosomal recruitment and translational control’, *Molecular Microbiology*, 43(1), pp. 239–246. doi: 10.1046/j.1365-2958.2002.02739.x.
- Molodtsov, V., Anikin, M. and McAllister, W. T. (2014) ‘The Presence of an RNA:DNA Hybrid That Is Prone to Slippage Promotes Termination by T7 RNA Polymerase’, *Journal of Molecular Biology*. Academic Press, 426(18), pp. 3095–3107. doi: 10.1016/J.JMB.2014.06.012.
- Monachino, E. *et al.* (2018) ‘A Primase-Induced Conformational Switch Controls the Stability of the Bacterial Replisome’, *bioRxiv*. Cold Spring Harbor Laboratory, p. 469312. doi: 10.1101/469312.
- Montange, R. K. and Batey, R. T. (2008) ‘Riboswitches: Emerging Themes in RNA Structure and Function’, *Annual Review of Biophysics*, 37(1), pp. 117–133. doi:

10.1146/annurev.biophys.37.032807.130000.

Morales-Filloy, H. G. *et al.* (2019) 'The 5'-NAD cap of RNAIII modulates toxin production in *Staphylococcus aureus* isolates', *bioRxiv*. Cold Spring Harbor Laboratory, p. 778233. doi: 10.1101/778233.

Morozov, Y. I. and Temiakov, D. (2016) 'Human mitochondrial transcription initiation complexes have similar topology on the light and heavy strand promoters', *Journal of Biological Chemistry*, 291(26), pp. 13432–13435. doi: 10.1074/jbc.C116.727966.

Mosaei, H. *et al.* (2018) 'Mode of Action of Kanglemycin A , an Ansamycin Natural Product that Is Active against Rifampicin- Resistant *Mycobacterium tuberculosis* Article Mode of Action of Kanglemycin A , an Ansamycin Natural Product that Is Active against Rifampicin-Resistant Mycob', *Molecular Cell*. Elsevier Inc., pp. 1–12. doi: 10.1016/j.molcel.2018.08.028.

Mukundan, V. T. and Phan, A. T. (2013) 'Bulges in G-quadruplexes: Broadening the definition of G-quadruplex-forming sequences', *Journal of the American Chemical Society*, 135(13), pp. 5017–5028. doi: 10.1021/ja310251r.

Mulder, M. A. *et al.* (1997) 'Mycobacterial promoters', *Tubercle and Lung Disease*, 78(6), pp. 211–223.

Murante, R. S., Henricksen, L. A. and Bambara, R. A. (1998) 'Junction ribonuclease: an activity in Okazaki fragment processing.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 95(5), pp. 2244–9. doi: 10.1073/pnas.95.5.2244.

Muzi-Falconi, M. *et al.* (2003) 'The DNA Polymerase γ -Primase Complex: Multiple Functions and Interactions', *The Scientific World JOURNAL*, 3, pp. 21–33. doi: 10.1100/tsw.2003.05.

Naue, N. *et al.* (2013) 'The helicase-binding domain of *Escherichia coli* DnaG primase interacts with the highly conserved C-terminal region of single-stranded DNA-binding protein.', *Nucleic acids research*. Oxford University Press, 41(8), pp. 4507–17. doi: 10.1093/nar/gkt107.

NEB (2015) 'NEBuilder HiFi DNA Assembly Cloning Kit', *Manual*. Available at: [https://international.neb.com/products/e5520-nebuilder-hifi-dna-assembly-cloning-kit#Product Information](https://international.neb.com/products/e5520-nebuilder-hifi-dna-assembly-cloning-kit#Product%20Information) (Accessed: 30 October 2019).

Nechaev, S. *et al.* (2002) 'A novel bacteriophage-encoded RNA polymerase binding protein inhibits transcription initiation and abolishes transcription termination by host RNA polymerase', *Journal of Molecular Biology*. Academic Press, 320(1), pp. 11–22. doi: 10.1016/S0022-2836(02)00420-5.

Nechaev, S. and Severinov, K. (1999) 'Inhibition of *Escherichia coli* RNA polymerase by bacteriophage T7 gene 2 protein', *Journal of Molecular Biology*. Academic Press, 289(4), pp. 815–826. doi: 10.1006/jmbi.1999.2782.

Nechaev, S. and Severinov, K. (2003) 'Bacteriophage-Induced Modifications of Host RNA Polymerase', *Annual Review of Microbiology*, 57(1), pp. 301–322. doi: 10.1146/annurev.micro.57.030502.090942.

Negrete, A. and Shiloach, J. (2015) 'Constitutive expression of the sRNA GadY decreases acetate production and improves *E. coli* growth', *Microbial Cell Factories*. BioMed Central,

14(1), pp. 1–10. doi: 10.1186/s12934-015-0334-1.

Nelson, D., Loomis, L. and Fischetti, V. A. (2001) ‘Prevention and elimination of upper respiratory colonization of mice by group A streptococci by using a bacteriophage lytic enzyme’, *Proceedings of the National Academy of Sciences*, 98(7), pp. 4107–4112. doi: 10.1073/pnas.061038398.

Nesbit, C. E. *et al.* (1995) ‘Transcriptional regulation of repressor synthesis in mycobacteriophage L5’, *Molecular Microbiology*, 17(6), pp. 1045–1056. doi: 10.1111/j.1365-2958.1995.mmi_17061045.x.

NewEngland Biolabs (2015) *T7 Express Competent E. coli (High Efficiency)*. Available at: [https://international.neb.com/products/c2566-t7-express-competent-e-coli-high-efficiency#Product Information](https://international.neb.com/products/c2566-t7-express-competent-e-coli-high-efficiency#Product%20Information) (Accessed: 30 October 2019).

NewEngland Biolabs (2018) *High Efficiency Transformation (C2987H/C2987I) | NEB*. doi: dx.doi.org/10.17504/protocols.io.isxcefn.

Newton-Foot, M. and Gey Van Pittius, N. C. (2013) ‘The complex architecture of mycobacterial promoters’, *Tuberculosis*, pp. 60–74. doi: 10.1016/j.tube.2012.08.003.

Nguta, J. M. *et al.* (2015) ‘Medicinal plants used to treat TB in Ghana’, *International Journal of Mycobacteriology*. No longer published by Elsevier, 4(2), pp. 116–123. doi: 10.1016/J.IJMYCO.2015.02.003.

Nieth, A. *et al.* (2015) ‘A first step toward liposome-mediated intracellular bacteriophage therapy’, *Expert Opinion on Drug Delivery*, 12(9), pp. 1411–1424. doi: 10.1517/17425247.2015.1043125.

Nowotny, M. *et al.* (2005) ‘Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis.’, *Cell*. Elsevier, 121(7), pp. 1005–16. doi: 10.1016/j.cell.2005.04.024.

Nübel, G., Sorgenfrei, F. A. and Jäschke, A. (2017) ‘Boronate affinity electrophoresis for the purification and analysis of cofactor-modified RNAs’, *Methods*, 117, pp. 14–20. doi: 10.1016/j.ymeth.2016.09.008.

Nudler, E. *et al.* (1997) ‘The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase.’, *Cell*, 89(1), pp. 33–41. doi: 10.1016/s0092-8674(00)80180-4.

Nudler, E. (2009) ‘RNA Polymerase Active Center: The Molecular Engine of Transcription’, *Annual Review of Biochemistry*, 78(1), pp. 335–361. doi: 10.1146/annurev.biochem.76.052705.164655.

O’Brien, E. *et al.* (2017) ‘The [4Fe4S] cluster of human DNA primase functions as a redox switch using DNA charge transport’, *Science*, 355(6327), p. eaag1789. doi: 10.1126/science.aag1789.

O’Brien, E., Holt, M. E., *et al.* (2018) ‘Substrate Binding Regulates Redox Signaling in Human DNA Primase.’, *Journal of the American Chemical Society*. NIH Public Access, 140(49), pp. 17153–17162. doi: 10.1021/jacs.8b09914.

O’Brien, E., Salay, L. E., *et al.* (2018) ‘Yeast require redox switching in DNA primase.’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 115(52), pp. 13186–13191. doi: 10.1073/pnas.1810715115.

- O'Donnell, M., Langston, L. and Stillman, B. (2013) 'Principles and concepts of DNA replication in bacteria, archaea, and eukarya.', *Cold Spring Harbor perspectives in biology*. Cold Spring Harbor Laboratory Press, 5(7). doi: 10.1101/cshperspect.a010108.
- Ogami, K., Chen, Y. and Manley, J. L. (2018) 'RNA surveillance by the nuclear RNA exosome: mechanisms and significance.', *Non-coding RNA*. Multidisciplinary Digital Publishing Institute (MDPI), 4(1). doi: 10.3390/ncrna4010008.
- Ogawa, T. and Okazaki, T. (1984) 'Function of RNase H in DNA replication revealed by RNase H defective mutants of *Escherichia coli*', *MGG Molecular & General Genetics*. Springer-Verlag, 193(2), pp. 231–237. doi: 10.1007/BF00330673.
- Oldfield, L. M. and Hatfull, G. F. (2014) 'Mutational analysis of the mycobacteriophage BPs promoter PR reveals context-dependent sequences for mycobacterial gene expression.', *Journal of bacteriology*. American Society for Microbiology (ASM), 196(20), pp. 3589–97. doi: 10.1128/JB.01801-14.
- Ooi, W.-Y. *et al.* (2018) 'A *Thermus* phage protein inhibits host RNA polymerase by preventing template DNA strand loading during open promoter complex formation', *Nucleic Acids Research*. Oxford University Press, 46(1), pp. 431–441. doi: 10.1093/nar/gkx1162.
- Opdyke, J. A. *et al.* (2011) 'RNase III participates in *gadY*-dependent cleavage of the *gadX-gadW* mRNA', *Journal of Molecular Biology*. Elsevier B.V., 406(1), pp. 29–43. doi: 10.1016/j.jmb.2010.12.009.
- Orsini, G. *et al.* (1993) 'The *asiA* gene of bacteriophage T4 codes for the anti-sigma 70 protein.', *Journal of bacteriology*. American Society for Microbiology (ASM), 175(1), pp. 85–93. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8416914> (Accessed: 5 November 2018).
- Osmundson, J. and Darst, S. A. (2013) 'Biochemical insights into the function of phage G1 gp67 in *Staphylococcus aureus*.', *Bacteriophage*. Taylor & Francis, 3(1), p. e24767. doi: 10.4161/bact.24767.
- Ouhammouch, M. *et al.* (1995) 'Bacteriophage T4 MotA and AsiA proteins suffice to direct *Escherichia coli* RNA polymerase to initiate transcription at T4 middle promoters.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 92(5), pp. 1451–5. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7877999> (Accessed: 5 November 2018).
- Paget, M. (2015) 'Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution', *Biomolecules*, 5(3), pp. 1245–1265. doi: 10.3390/biom5031245.
- Paget, M. S. (2015) 'Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution.', *Biomolecules*. Multidisciplinary Digital Publishing Institute (MDPI), 5(3), pp. 1245–65. doi: 10.3390/biom5031245.
- Pai, M. *et al.* (2016) 'Tuberculosis', *Nature Reviews Disease Primers*. doi: 10.1038/nrdp.2016.76.
- Pal, D. *et al.* (2003) 'Analysis of Regions within the Bacteriophage T4 AsiA Protein Involved in its Binding to the $\sigma 70$ Subunit of *E. coli* RNA Polymerase and its Role as a Transcriptional Inhibitor and Co-activator', *Journal of Molecular Biology*. Academic Press, 325(5), pp. 827–841. doi: 10.1016/S0022-2836(02)01307-4.
- Parish, T. and Roberts, D. M. (eds) (2015) *Mycobacteria Protocols*. New York, NY: Springer

New York (Methods in Molecular Biology). doi: 10.1007/978-1-4939-2450-9.

Paul, B. J., Barker, M. M., *et al.* (2004) 'DksA: A Critical Component of the Transcription Initiation Machinery that Potentiates the Regulation of rRNA Promoters by ppGpp and the Initiating NTP', *Cell*. Cell Press, 118(3), pp. 311–322. doi: 10.1016/J.CELL.2004.07.009.

Paul, B. J., Ross, W., *et al.* (2004) 'rRNA Transcription in *Escherichia coli*', *Annual Review of Genetics*. Annual Reviews, 38(1), pp. 749–770. doi: 10.1146/annurev.genet.38.072902.091347.

Pellegrini, L. (2012) 'The Pol α -Primase Complex', in *Sub-cellular biochemistry*, pp. 157–169. doi: 10.1007/978-94-007-4572-8_9.

Peña, C. E. A. *et al.* (1997) 'Characterization of the Mycobacteriophage L5 attachment site, attP', *Journal of Molecular Biology*, 266(1), pp. 76–92. doi: 10.1006/jmbi.1996.0774.

Perumal, A. S. *et al.* (2018) 'RbpA relaxes promoter selectivity of *M. tuberculosis* RNA polymerase', *Nucleic Acids Research*. Oxford University Press, 46(19), pp. 10106–10118. doi: 10.1093/nar/gky714.

Peters, J. M. *et al.* (2012) 'Rho and NusG suppress pervasive antisense transcription in *Escherichia coli*.', *Genes & development*. Cold Spring Harbor Laboratory Press, 26(23), pp. 2621–33. doi: 10.1101/gad.196741.112.

Pope, W. H. *et al.* (2011) 'Cluster k mycobacteriophages: Insights into the evolutionary origins of mycobacteriophage tm4', *PLoS ONE*. doi: 10.1371/journal.pone.0026750.

Postow, L. *et al.* (2001) 'Topological challenges to DNA replication: conformations at the fork.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 98(15), pp. 8219–26. doi: 10.1073/pnas.111006998.

Potrykus, K. *et al.* (2011) 'ppGpp is the major source of growth rate control in *E. coli*', *Environmental Microbiology*. John Wiley & Sons, Ltd (10.1111), 13(3), pp. 563–575. doi: 10.1111/j.1462-2920.2010.02357.x.

Price, A. (1992) 'Action of *Escherichia coli* and human 5' \rightarrow 3' exonuclease functions at incised apurinic/apyrimidinic sites in DNA', *FEBS Letters*. John Wiley & Sons, Ltd, 300(1), pp. 101–104. doi: 10.1016/0014-5793(92)80173-E.

Price, A. and Lindahl, T. (1991) 'Enzymic release of 5'-terminal deoxyribose phosphate residues from damaged DNA in human cells', *Biochemistry*, 30(35), pp. 8631–8637. doi: 10.1021/bi00099a020.

Provvedi, R. *et al.* (2008) 'SigF controls carotenoid pigment production and affects transformation efficiency and hydrogen peroxide sensitivity in *Mycobacterium smegmatis*', *Journal of Bacteriology*. doi: 10.1128/JB.00714-08.

Puiu, M. and Julius, C. (2019) 'Bacteriophage gene products as potential antimicrobials against tuberculosis.', *Biochemical Society transactions*. Portland Press Limited, 47(3), pp. 847–860. doi: 10.1042/BST20180506.

Pulvermacher, S. C., Stauffer, L. T. and Stauffer, G. V. (2009) 'Role of the sRNA GcvB in regulation of *cycA* in *Escherichia coli*', *Microbiology*, 155(1), pp. 106–114. doi: 10.1099/mic.0.023598-0.

Pupov, D. *et al.* (2014) 'Distinct functions of the RNA polymerase σ subunit region 3.2 in RNA priming and promoter escape', *Nucleic Acids Research*, 42(7), pp. 4494–4504. doi:

10.1093/nar/gkt1384.

QIAGEN (2012) 'QIAprep Spin Miniprep Kit'. Available at: https://molbio.mgh.harvard.edu/szostakweb/protocols/clone_seq/qpmini_spin.pdf (Accessed: 30 October 2019).

Raghunathan, N. *et al.* (2018) 'Genome-wide relationship between R-loop formation and antisense transcription in *Escherichia coli*', *Nucleic Acids Research*. Narnia, 46(7), pp. 3400–3411. doi: 10.1093/nar/gky118.

Ramanathan, A., Robb, G. B. and Chan, S.-H. (2016) 'mRNA capping: biological functions and applications.', *Nucleic acids research*. England, 44(16), pp. 7511–7526. doi: 10.1093/nar/gkw551.

Régnier, P. and Marujo, P. E. (2013) 'Polyadenylation and Degradation of RNA in Prokaryotes'. Landes Bioscience. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK6253/> (Accessed: 24 January 2020).

Reichenbach, B. *et al.* (2008) 'The small RNA GlmY acts upstream of the sRNA GlmZ in the activation of glmS expression and is subject to regulation by polyadenylation in *Escherichia coli*.', *Nucleic acids research*. England, 36(8), pp. 2570–2580. doi: 10.1093/nar/gkn091.

Repoila, F. and Darfeuille, F. (2009) 'Small regulatory non-coding RNAs in bacteria: physiology and mechanistic aspects', *Biology of the Cell*, 101(2), pp. 117–131. doi: 10.1042/BC20070137.

Revyakin, A. *et al.* (2006) 'Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching.', *Science (New York, N.Y.)*. NIH Public Access, 314(5802), pp. 1139–43. doi: 10.1126/science.1131398.

Richardson, J. P. (1982) 'Activation of rho protein ATPase requires simultaneous interaction at two kinds of nucleic acid-binding sites.', *The Journal of biological chemistry*, 257(10), pp. 5760–6. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/6175630> (Accessed: 6 September 2019).

Roach, D. R. *et al.* (2017) 'Synergy between the Host Immune System and Bacteriophage Is Essential for Successful Phage Therapy against an Acute Respiratory Pathogen', *Cell Host & Microbe*. Cell Press, 22(1), pp. 38–47.e4. doi: 10.1016/J.CHOM.2017.06.018.

Roberts, J. and Park, J.-S. (2004) 'Mfd, the bacterial transcription repair coupling factor: translocation, repair and termination', *Current Opinion in Microbiology*. Elsevier Current Trends, 7(2), pp. 120–125. doi: 10.1016/J.MIB.2004.02.014.

Roberts, J. W. (1969) 'Termination factor for RNA synthesis', *Nature*. Nature Publishing Group, 224(5225), pp. 1168–1174. doi: 10.1038/2241168a0.

Roberts, J. W. (2019) 'Mechanisms of Bacterial Transcription Termination', *Journal of Molecular Biology*. Academic Press, pp. 4030–4039. doi: 10.1016/j.jmb.2019.04.003.

Robins, P. *et al.* (1994) 'Structural and functional homology between mammalian DNase IV and the 5'-nuclease domain of *Escherichia coli* DNA polymerase I.', *The Journal of biological chemistry*, 269(46), pp. 28535–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7961795> (Accessed: 2 October 2019).

Rodina, A. and Godson, G. N. (2006) 'Role of conserved amino acids in the catalytic activity of *Escherichia coli* primase.', *Journal of bacteriology*. American Society for Microbiology

Journals, 188(10), pp. 3614–21. doi: 10.1128/JB.188.10.3614-3621.2006.

Rossi, M. L. and Bambara, R. A. (2006) 'Reconstituted Okazaki fragment processing indicates two pathways of primer removal', *Journal of Biological Chemistry*, 281(36), pp. 26051–26061. doi: 10.1074/jbc.M604805200.

Rouault, T. A. (2015) 'Mammalian iron–sulphur proteins: novel insights into biogenesis and function', *Nature Reviews Molecular Cell Biology*, 16(1), pp. 45–55. doi: 10.1038/nrm3909.

Roucourt, B. and Lavigne, R. (2009) 'The role of interactions between phage and bacterial proteins within the infected cell: a diverse and puzzling interactome', *Environmental Microbiology*. John Wiley & Sons, Ltd (10.1111), 11(11), pp. 2789–2805. doi: 10.1111/j.1462-2920.2009.02029.x.

Rudolph, C. J. *et al.* (2013) 'Avoiding chromosome pathology when replication forks collide', *Nature*. Europe PMC Funders, 500(7464), pp. 608–11. doi: 10.1038/NATURE12312.

Ryan, E. M. *et al.* (2012) 'Synergistic phage-antibiotic combinations for the control of *Escherichia coli* biofilms *in vitro*', *FEMS Immunology & Medical Microbiology*. Oxford University Press, 65(2), pp. 395–398. doi: 10.1111/j.1574-695X.2012.00977.x.

Rybniiker, J. *et al.* (2008) 'Identification of three cytotoxic early proteins of mycobacteriophage L5 leading to growth inhibition in *Mycobacterium smegmatis*', *Microbiology*. Microbiology Society, 154(8), pp. 2304–2314. doi: 10.1099/mic.0.2008/017004-0.

Rymer, R. U. *et al.* (2012) 'Binding mechanism of metal-NTP substrates and stringent-response alarmones to bacterial DnaG-type primases', *Structure*, 20(9), pp. 1478–1489. doi: 10.1016/j.str.2012.05.017.

Sachs, A. B. (1993) 'Messenger RNA degradation in eukaryotes', *Cell*. Cell Press, 74(3), pp. 413–421. doi: 10.1016/0092-8674(93)80043-E.

Santangelo, T. J. and Artsimovitch, I. (2011) 'Termination and antitermination: RNA polymerase runs a stop sign.', *Nature reviews. Microbiology*. NIH Public Access, 9(5), pp. 319–29. doi: 10.1038/nrmicro2560.

São-José, C. (2018) 'Engineering of Phage-Derived Lytic Enzymes: Improving Their Potential as Antimicrobials.', *Antibiotics (Basel, Switzerland)*. Multidisciplinary Digital Publishing Institute (MDPI), 7(2). doi: 10.3390/antibiotics7020029.

Sarkar, N. (1997) 'Polyadenylation of mRNA in prokaryotes.', *Annual review of biochemistry*. United States, 66, pp. 173–197. doi: 10.1146/annurev.biochem.66.1.173.

Sauguet, L. *et al.* (2010) 'Shared active site architecture between the large subunit of eukaryotic primase and DNA photolyase.', *PloS one*. Public Library of Science, 5(4), p. e10083. doi: 10.1371/journal.pone.0010083.

Schmelcher, M., Donovan, D. M. and Loessner, M. J. (2012) 'Bacteriophage endolysins as novel antimicrobials', *Future Microbiology*, pp. 1147–1171. doi: 10.2217/fmb.12.97.

Schneider, A. *et al.* (1998) 'Primase Activity of Human DNA Polymerase -Primase: divalent cations stabilize the enzyme activity of the p48 subunit', *Journal of Biological Chemistry*, 273(34), pp. 21608–21615. doi: 10.1074/jbc.273.34.21608.

Schuergers, N. *et al.* (2014) 'Binding of the RNA chaperone Hfq to the type IV pilus base is crucial for its function in *S ynechocystis* sp. PCC 6803', *Molecular Microbiology*, 92(4), pp.

840–852. doi: 10.1111/mmi.12595.

Schutz, C. *et al.* (2010) ‘Clinical management of tuberculosis and HIV-1 co-infection.’, *The European respiratory journal*. European Respiratory Society, 36(6), pp. 1460–81. doi: 10.1183/09031936.00110210.

Selby, C. and Sancar, A. (1993) ‘Molecular mechanism of transcription-repair coupling’, *Science*, 260(5104), pp. 53–58. doi: 10.1126/science.8465200.

Semenya, S. S. and Maroyi, A. (2013) ‘Medicinal plants used for the treatment of tuberculosis by Bapedi traditional healers in three districts of the Limpopo Province, South Africa.’, *African journal of traditional, complementary, and alternative medicines : AJTCAM*. African Traditional Herbal Medicine Supporters Initiative, 10(2), pp. 316–23. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24146456> (Accessed: 7 January 2019).

Sharifi-Rad, J. *et al.* (2017) ‘Medicinal plants used in the treatment of tuberculosis - Ethnobotanical and ethnopharmacological approaches’, *Biotechnology Advances*. Elsevier. doi: 10.1016/j.BIOTECHADV.2017.07.001.

Sheaff, R. J. and Kuchta, R. D. (1993) ‘Mechanism of calf thymus DNA primase: Slow initiation, rapid polymerization, and intelligent termination’, *Biochemistry*. American Chemical Society, 32(12), pp. 3027–3037. doi: 10.1021/bi00063a014.

Shell, S. S. *et al.* (2015) ‘Leaderless Transcripts and Small Proteins Are Common Features of the Mycobacterial Translational Landscape’, *PLoS Genetics*, 11(11), pp. 1–31. doi: 10.1371/journal.pgen.1005641.

Siddiqui, K., On, K. F. and Diffley, J. F. X. (2013) ‘Regulating DNA Replication in Eukarya’, *Cold Spring Harbor Perspectives in Biology*, 5(9), pp. a012930–a012930. doi: 10.1101/cshperspect.a012930.

Silva, I. J. *et al.* (2011) ‘Importance and key events of prokaryotic RNA decay: The ultimate fate of an RNA molecule’, *Wiley Interdisciplinary Reviews: RNA*, pp. 818–836. doi: 10.1002/wrna.94.

Silva Miranda, M. *et al.* (2012) ‘The Tuberculous Granuloma: An Unsuccessful Host Defence Mechanism Providing a Safety Shelter for the Bacteria?’, *Clinical and Developmental Immunology*, 2012, pp. 1–14. doi: 10.1155/2012/139127.

Simonetta, K. R. *et al.* (2009) ‘The mechanism of ATP-dependent primer-template recognition by a clamp loader complex.’, *Cell*. NIH Public Access, 137(4), pp. 659–71. doi: 10.1016/j.cell.2009.03.044.

Singla, S. *et al.* (2016) ‘Encapsulation of Bacteriophage in Liposome Accentuates Its Entry in to Macrophage and Shields It from Neutralizing Antibodies’, *PLOS ONE*. Edited by J. A. Bengoechea. Public Library of Science, 11(4), p. e0153777. doi: 10.1371/journal.pone.0153777.

Skarstad, K. and Boye, E. (1994) ‘The initiator protein DnaA: evolution, properties and function’, *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 1217(2), pp. 111–130. doi: 10.1016/0167-4781(94)90025-6.

Smith, I. (2003) ‘*Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence.’, *Clinical microbiology reviews*. American Society for Microbiology (ASM), 16(3), pp. 463–96. doi: 10.1128/CMR.16.3.463-496.2003.

- Smith, T., Wolff, K. A. and Nguyen, L. (2013) 'Molecular biology of drug resistance in *Mycobacterium tuberculosis*.', *Current topics in microbiology and immunology*. NIH Public Access, 374, pp. 53–80. doi: 10.1007/82_2012_279.
- Sosunov, V. *et al.* (2003) 'Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase.', *The EMBO journal*. European Molecular Biology Organization, 22(9), pp. 2234–44. doi: 10.1093/emboj/cdg193.
- Stamford, N. P. J. *et al.* (1992) 'Enriched sources of *Escherichia coli* replication proteins: The dnaG primase is a zinc metalloprotein', *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*. Elsevier, 1132(1), pp. 17–25. doi: 10.1016/0167-4781(92)90047-4.
- Stauffer, L. T. and Stauffer, G. V. (2012) 'The *Escherichia coli* GcvB sRNA Uses Genetic Redundancy to Control *cycA* Expression', *ISRN Microbiology*, 2012(c), pp. 1–10. doi: 10.5402/2012/636273.
- Steitz, T. A. and Steitz, J. A. (1993) 'A general two-metal-ion mechanism for catalytic RNA.', *Proceedings of the National Academy of Sciences*, 90(14), pp. 6498–6502. doi: 10.1073/pnas.90.14.6498.
- Stewart, J. A. *et al.* (2008) 'Dynamic removal of replication protein A by Dna2 facilitates primer cleavage during Okazaki fragment processing in *Saccharomyces cerevisiae*.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 283(46), pp. 31356–65. doi: 10.1074/jbc.M805965200.
- Stoffels, K. *et al.* (2013) 'From Multidrug- to Extensively Drug-Resistant Tuberculosis: Upward Trends as Seen from a 15-Year Nationwide Study', *PLoS ONE*. Edited by J. L. Herrmann. Public Library of Science, 8(5), p. e63128. doi: 10.1371/journal.pone.0063128.
- Stratagene (2002) 'QuikChange® II XL Site-Directed Mutagenesis Kit', *Mutagenesis*, 200521(5), pp. 1158–1165. Available at: <http://www.genomics.agilent.com/files/Manual/200521.pdf>.
- Sula, L., Sulová, J. and Stolcpartová, M. (1981) 'Therapy of experimental tuberculosis in guinea pigs with mycobacterial phages DS-6A, GR-21 T, My-327.', *Czechoslovak medicine*, 4(4), pp. 209–14. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7327068> (Accessed: 22 October 2018).
- Sulakvelidze, A., Alavidze, Z. and Morris, J. G. (2001) 'MINIREVIEW Bacteriophage Therapy', 45(3), pp. 649–659. doi: 10.1128/AAC.45.3.649–659.2001.
- Sulej, A. A. *et al.* (2012) 'Sequence-specific cleavage of the RNA strand in DNA-RNA hybrids by the fusion of ribonuclease H with a zinc finger', *Nucleic Acids Research*, 40(22), pp. 11563–11570. doi: 10.1093/nar/gks885.
- Sun, F. *et al.* (2012) 'Biochemical Issues in Estimation of Cytosolic Free NAD/NADH Ratio', *PLoS ONE*. Edited by L. Song. Public Library of Science, 7(5), p. e34525. doi: 10.1371/journal.pone.0034525.
- Sun, W., Schoneich, J. and Godson, G. N. (1999) 'A mutant *Escherichia coli* primase defective in elongation of primer RNA chains.', *Journal of bacteriology*. American Society for Microbiology Journals, 181(12), pp. 3761–7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10368151> (Accessed: 3 October 2019).
- Suzuki, M. *et al.* (1993) 'RNA priming coupled with DNA synthesis on natural template by calf thymus DNA polymerase .alpha.-primase', *Biochemistry*, 32(47), pp. 12782–12792. doi:

10.1021/bi00210a030.

Swart, J. R. and Griep, M. A. (1993) 'Primase from *Escherichia coli* primes single-stranded templates in the absence of single-stranded DNA-binding protein or other auxiliary proteins. Template sequence requirements based on the bacteriophage G4 complementary strand origin and Okazaki fragment initiation sites.', *The Journal of biological chemistry*, 268(17), pp. 12970–6. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8509429> (Accessed: 3 July 2019).

Swart, J. R. and Griep, M. A. (1995) 'Primer synthesis kinetics by *Escherichia coli* primase on single-stranded DNA templates', *Biochemistry*. American Chemical Society, 34(49), pp. 16097–16106. doi: 10.1021/bi00049a025.

Taanman, J. W. (1999) 'The mitochondrial genome: Structure, transcription, translation and replication', *Biochimica et Biophysica Acta - Bioenergetics*, 1410(2), pp. 103–123. doi: 10.1016/S0005-2728(98)00161-3.

Tabib-Salazar, A. *et al.* (2013) 'The actinobacterial transcription factor RbpA binds to the principal sigma subunit of RNA polymerase', *Nucleic Acids Research*. Oxford University Press, 41(11), pp. 5679–5691. doi: 10.1093/nar/gkt277.

Tagami, S. *et al.* (2014) 'Structural basis for promoter specificity switching of RNA polymerase by a phage factor.', *Genes & development*. Cold Spring Harbor Laboratory Press, 28(5), pp. 521–31. doi: 10.1101/gad.233916.113.

Tamm, J. and Polisky, B. (1985) 'Characterization of the ColE1 primer-RNA1 complex: analysis of a domain of ColE1 RNA1 necessary for its interaction with primer RNA.', *Proceedings of the National Academy of Sciences of the United States of America*, 82(8), pp. 2257–2261. doi: 10.1073/pnas.82.8.2257.

Temiaikov, D. *et al.* (2005) 'Structural Basis of Transcription Inhibition by Antibiotic Streptolydigin', *Molecular Cell*. Cell Press, 19(5), pp. 655–666. doi: 10.1016/J.MOLCEL.2005.07.020.

Teng, T., Liu, J. and Wei, H. (2015) 'Anti-mycobacterial peptides: from human to phage.', *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology*. Karger Publishers, 35(2), pp. 452–66. doi: 10.1159/000369711.

Thomsen, N. D. *et al.* (2016) 'Molecular mechanisms of substrate-controlled ring dynamics and substepping in a nucleic acid-dependent hexameric motor.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 113(48), pp. E7691–E7700. doi: 10.1073/pnas.1616745113.

Thrombin Cleavage Capture Kit (no date). Available at: http://wolfson.huji.ac.il/purification/PDF/Protease_fusion_cleavage/NOVAGEN_Thrombin_kit.pdf (Accessed: 30 October 2019).

Tiberi, S. *et al.* (2018) 'New drugs and perspectives for new anti-tuberculosis regimens', *Pulmonology*. Elsevier Doyma, 24(2), pp. 86–98. doi: 10.1016/J.RPPNEN.2017.10.009.

Tkhilaishvili, T. *et al.* (2018) 'Bacteriophage Sb-1 enhances antibiotic activity against biofilm, degrades exopolysaccharide matrix and targets persisters of *Staphylococcus aureus*', *International Journal of Antimicrobial Agents*. Elsevier, 52(6), pp. 842–853. doi: 10.1016/J.IJANTIMICAG.2018.09.006.

- Tougu, K., Peng, H. and Mariani, K. J. (1994) 'Identification of a domain of *Escherichia coli* primase required for functional interaction with the DnaB helicase at the replication fork.', *The Journal of biological chemistry*, 269(6), pp. 4675–82. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8308039> (Accessed: 24 September 2019).
- Touloukianov, I., Artsimovitch, I. and Landick, R. (2001) 'Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 292(5517), pp. 730–3. doi: 10.1126/science.1057738.
- Traxler, M. F. *et al.* (2008) 'The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*.', *Molecular microbiology*. NIH Public Access, 68(5), pp. 1128–48. doi: 10.1111/j.1365-2958.2008.06229.x.
- Trigo, G. *et al.* (2013) 'Phage Therapy Is Effective against Infection by *Mycobacterium ulcerans* in a Murine Footpad Model', *PLoS Neglected Tropical Diseases*. Edited by C. Johnson. Public Library of Science, 7(4), p. e2183. doi: 10.1371/journal.pntd.0002183.
- Tse, E. C. M., Zwang, T. J. and Barton, J. K. (2017) 'The Oxidation State of [4Fe4S] Clusters Modulates the DNA-Binding Affinity of DNA Repair Proteins', *Journal of the American Chemical Society*. American Chemical Society, 139(36), pp. 12784–12792. doi: 10.1021/jacs.7b07230.
- Turchi, J. J. *et al.* (1994) 'Enzymatic completion of mammalian lagging-strand DNA replication.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 91(21), pp. 9803–7. doi: 10.1073/pnas.91.21.9803.
- Turk, E. M. *et al.* (2013) 'The Mitochondrial RNA Landscape of *Saccharomyces cerevisiae*', *PLoS ONE*. Edited by S. Granneman. Public Library of Science, 8(10), p. e78105. doi: 10.1371/journal.pone.0078105.
- Udwadia, Z. and Vendoti, D. (2013) 'Totally drug-resistant tuberculosis (TDR-TB) in India: every dark cloud has a silver lining.', *Journal of epidemiology and community health*. BMJ Publishing Group Ltd, 67(6), pp. 471–2. doi: 10.1136/jech-2012-201640.
- Vaithiyalingam, S. *et al.* (2014) 'Insights into Eukaryotic Primer Synthesis from Structures of the p48 Subunit of Human DNA Primase', *Journal of Molecular Biology*, 426(3), pp. 558–569. doi: 10.1016/j.jmb.2013.11.007.
- Varik, V. *et al.* (2017) 'HPLC-based quantification of bacterial housekeeping nucleotides and alarmone messengers ppGpp and pppGpp', *Scientific Reports*. Nature Publishing Group, 7(1), p. 11022. doi: 10.1038/s41598-017-10988-6.
- Vasilyev, N. and Serganov, A. (2015) 'Structures of RNA complexes with the *Escherichia coli* RNA pyrophosphohydrolase RppH Unveil the basis for specific 5'-End-dependent mRNA decay', *Journal of Biological Chemistry*, 290(15), pp. 9487–9499. doi: 10.1074/jbc.M114.634824.
- Vassilyev, D. G. *et al.* (2002) 'Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution', *Nature*. Nature Publishing Group, 417(6890), pp. 712–719. doi: 10.1038/nature752.
- Vassilyev, D. G., Vassilyeva, M. N., Zhang, J., *et al.* (2007) 'Structural basis for substrate loading in bacterial RNA polymerase', *Nature*, 448(7150), pp. 163–168. doi: 10.1038/nature05931.

Vassilyev, D. G., Vassilyeva, M. N., Perederina, A., *et al.* (2007) ‘Structural basis for transcription elongation by bacterial RNA polymerase’, *Nature*, 448(7150), pp. 157–162. doi: 10.1038/nature05932.

Vaudry, W. (2003) “‘To BCG or not to BCG, that is the question!’”. The challenge of BCG vaccination: Why can’t we get it right?, *Paediatrics and Child Health*. Oxford University Press, pp. 141–144. doi: 10.1093/pch/8.3.141.

Vvedenskaya, I. O. *et al.* (2018) ‘CapZyme-Seq Comprehensively Defines Promoter-Sequence Determinants for RNA Capping with NAD⁺’, *Molecular Cell*. Elsevier, 70(3), pp. 553–564.e9. doi: 10.1016/j.molcel.2018.03.014.

Waagmeester, A., Thompson, J. and Reyrat, J.-M. (2005) ‘Identifying sigma factors in *Mycobacterium smegmatis* by comparative genomic analysis’, *Trends in Microbiology*. Elsevier Current Trends, 13(11), pp. 505–509. doi: 10.1016/J.TIM.2005.08.009.

Wall, R. J. *et al.* (2005) ‘Genetically enhanced cows resist intramammary *Staphylococcus aureus* infection’, *Nature Biotechnology*, 23(4), pp. 445–451. doi: 10.1038/nbt1078.

Walters, R. W. *et al.* (2017) ‘Identification of NAD⁺-capped mRNAs in *Saccharomyces cerevisiae*’, *Proceedings of the National Academy of Sciences*, 114(3), pp. 480 LP – 485. Available at: <http://www.pnas.org/content/114/3/480.abstract>.

Wang, J. *et al.* (2019) ‘Quantifying the RNA cap epitranscriptome reveals novel caps in cellular and viral RNA’, *bioRxiv*. Cold Spring Harbor Laboratory, p. 683045. doi: 10.1101/683045.

Wang, X. *et al.* (2017) ‘Engineering *Escherichia coli* Nicotinic Acid Mononucleotide Adenylyltransferase for Fully Active Amidated NAD Biosynthesis.’, *Applied and environmental microbiology*. American Society for Microbiology (ASM), 83(13). doi: 10.1128/AEM.00692-17.

Wang, Y. *et al.* (2019) ‘NAD⁺-capped RNAs are widespread in the Arabidopsis transcriptome and can probably be translated’, *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 116(24), pp. 12094–12102. doi: 10.1073/PNAS.1903682116.

Wang, Z. *et al.* (2018) ‘A phage lysin fused to a cell-penetrating peptide kills intracellular Methicillin-resistant *Staphylococcus aureus* in keratinocytes and has potential as a treatment for skin infections in mice’, *Applied and Environmental Microbiology*.

Wanrooij, P. H. *et al.* (2012) ‘A hybrid G-quadruplex structure formed between RNA and DNA explains the extraordinary stability of the mitochondrial R-loop’, *Nucleic Acids Research*, 40(20), pp. 10334–10344. doi: 10.1093/nar/gks802.

Wanrooij, S. *et al.* (2008) ‘Human mitochondrial RNA polymerase primes lagging-strand DNA synthesis *in vitro*’, *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 105(32), pp. 11122–11127. doi: 10.1073/PNAS.0805399105.

Waters, L. S. and Storz, G. (2009) ‘Regulatory RNAs in Bacteria’, *Cell*, pp. 615–628. doi: 10.1016/j.cell.2009.01.043.

Wei, L. *et al.* (2013) ‘A mycobacteriophage-derived trehalose-6,6'-dimycolate-binding peptide containing both antimycobacterial and anti-inflammatory abilities’, *The FASEB Journal*. Federation of American Societies for Experimental Biology Bethesda, MD, USA, 27(8), pp. 3067–3077. doi: 10.1096/fj.13-227454.

- Weiner, B. E. *et al.* (2007) 'An iron-sulfur cluster in the C-terminal domain of the p58 subunit of human DNA primase.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 282(46), pp. 33444–51. doi: 10.1074/jbc.M705826200.
- Wendel, B. M., Courcelle, C. T. and Courcelle, J. (2014) 'Completion of DNA replication in *Escherichia coli*.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 111(46), pp. 16454–9. doi: 10.1073/pnas.1415025111.
- Westhof, E. and Auffinger, P. (2000) 'RNA Tertiary Structure', in *Encyclopedia of Analytical Chemistry*. Chichester, UK: John Wiley & Sons, Ltd. doi: 10.1002/9780470027318.a1428.
- White, H. B. (1976) 'Coenzymes as fossils of an earlier metabolic state', *Journal of Molecular Evolution*, 7(2), pp. 101–104. doi: 10.1007/BF01732468.
- Winston, C. A. and Mitruka, K. (2012) 'Treatment Duration for Patients with Drug-Resistant Tuberculosis, United States', *Emerging Infectious Diseases*, 18(7), pp. 1201–1202. doi: 10.3201/eid1807.120261.
- Winther, K. S., Roghanian, M. and Gerdes, K. (2018) 'Activation of the Stringent Response by Loading of RelA-tRNA Complexes at the Ribosomal A-Site.', *Molecular cell*. Elsevier, 70(1), pp. 95-105.e4. doi: 10.1016/j.molcel.2018.02.033.
- World Health Organization (2018) *Global tuberculosis report 2018 - Geneva, WHO report*. World Health Organization. doi: 10.1016/j.pharep.2017.02.021.
- Wright, A. *et al.* (2009) 'A controlled clinical trial of a therapeutic bacteriophage preparation in chronic otitis due to antibiotic-resistant *Pseudomonas aeruginosa* ; a preliminary report of efficacy', *Clinical Otolaryngology*. John Wiley & Sons, Ltd (10.1111), 34(4), pp. 349–357. doi: 10.1111/j.1749-4486.2009.01973.x.
- Wu, P. *et al.* (2017) 'The important conformational plasticity of DsrA sRNA for adapting multiple target regulation', *Nucleic Acids Research*. Oxford University Press, 45(16), pp. 9625–9639. doi: 10.1093/nar/gkx570.
- Wu, S. Y. and Platt, T. (1993) 'Transcriptional arrest of yeast RNA polymerase II by *Escherichia coli* Rho protein *in vitro*.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 90(14), pp. 6606–10. doi: 10.1073/pnas.90.14.6606.
- Xie, P. (2011) 'A model for dynamics of primer extension by eukaryotic DNA primase', *European Biophysics Journal*, 40(10), pp. 1157–1165. doi: 10.1007/s00249-011-0746-8.
- Xu, W. *et al.* (2006) 'Three new Nudix hydrolases from *Escherichia coli*.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 281(32), pp. 22794–8. doi: 10.1074/jbc.M603407200.
- Xu, Y. *et al.* (1997) 'Biochemical and mutational studies of the 5'-3' exonuclease of DNA polymerase I of *Escherichia coli*', *Journal of Molecular Biology*, 268(2), pp. 284–302. doi: 10.1006/jmbi.1997.0967.
- Xu, Y., Grindley, N. D. and Joyce, C. M. (2000) 'Coordination between the polymerase and 5'-nuclease components of DNA polymerase I of *Escherichia coli*.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 275(27), pp. 20949–55. doi: 10.1074/jbc.M909135199.

- Yan, J. *et al.* (2018) 'An archaeal primase functions as a nanoscale caliper to define primer length.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 115(26), pp. 6697–6702. doi: 10.1073/pnas.1806351115.
- Yang, Haiquan *et al.* (2014) 'Transcription regulation mechanisms of bacteriophages', *Bioengineered*. doi: 10.4161/bioe.32110.
- Yang, J. *et al.* (2005) 'The Oligomeric T4 Primase Is the Functional Form during Replication', *Journal of Biological Chemistry*, 280(27), pp. 25416–25423. doi: 10.1074/jbc.M501847200.
- Yang, W., Lee, J. Y. and Nowotny, M. (2006) 'Making and Breaking Nucleic Acids: Two-Mg²⁺-Ion Catalysis and Substrate Specificity', *Molecular Cell*. Cell Press, 22(1), pp. 5–13. doi: 10.1016/J.MOLCEL.2006.03.013.
- Yao, S., Sharp, J. S. and Bechhofer, D. H. (2009) 'Bacillus subtilis RNase J1 endonuclease and 5' exonuclease activities in the turnover of DeltaermC mRNA.', *RNA (New York, N.Y.)*. Cold Spring Harbor Laboratory Press, 15(12), pp. 2331–9. doi: 10.1261/rna.1749109.
- Yarnell, W. S. and Roberts, J. W. (1999) 'Mechanism of Intrinsic Transcription Termination and Antitermination', *Science*, 284(5414), pp. 611–615. doi: 10.1126/science.284.5414.611.
- Yoda, K. *et al.* (1988) 'RNA-primed initiation sites of DNA replication in the origin region of bacteriophage λ genome', *Nucleic Acids Research*. Narnia, 16(14), pp. 6531–6546. doi: 10.1093/nar/16.14.6531.
- Yoda, K. and Okazaki, T. (1991) 'Specificity of recognition sequence for *Escherichia coli* primase', *MGG Molecular & General Genetics*. Springer-Verlag, 227(1), pp. 1–8. doi: 10.1007/BF00260698.
- Yoshikawa, H. (1970) 'DNA synthesis and its regulation', *Tanpakushitsu kakusan koso. Protein, nucleic acid, enzyme*, pp. 325–335.
- Young, R., Wang, I.-N. and Roof, W. D. (2000) 'Phages will out: strategies of host cell lysis', *Trends in Microbiology*. Elsevier Current Trends, 8(3), pp. 120–128. doi: 10.1016/S0966-842X(00)01705-4.
- Yuzenkova, J. *et al.* (2003) 'Genome of *Xanthomonas oryzae* bacteriophage Xp10: an odd T-odd phage.', *Journal of molecular biology*, 330(4), pp. 735–748. doi: 10.1016/S0022-2836(03)00634-X.
- Yuzenkova, Y. *et al.* (2010) 'Stepwise mechanism for transcription fidelity', *BMC Biology*, 8(1), p. 54. doi: 10.1186/1741-7007-8-54.
- Yuzenkova, Y. *et al.* (2011) 'A new basal promoter element recognized by RNA polymerase core enzyme.', *The EMBO journal*, 30(18), pp. 3766–75. doi: 10.1038/emboj.2011.252.
- Yuzenkova, Y. and Zenkin, N. (2010) 'Central role of the RNA polymerase trigger loop in intrinsic RNA hydrolysis.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 107(24), pp. 10878–83. doi: 10.1073/pnas.0914424107.
- Yuzhakov, A., Kelman, Z. and O'Donnell, M. (1999) 'Trading places on DNA-A three-point switch underlies primer handoff from primase to the replicative DNA polymerase', *Cell*, 96(1), pp. 153–163. doi: 10.1016/S0092-8674(00)80968-X.
- Zeng, Q., McNally, R. R. and Sundin, G. W. (2013) 'Global small RNA chaperone Hfq and

- regulatory small RNAs are important virulence regulators in *Erwinia amylovora*.', *Journal of bacteriology*. American Society for Microbiology Journals, 195(8), pp. 1706–17. doi: 10.1128/JB.02056-12.
- Zenkin, N. *et al.* (2006) 'The mechanism of DNA replication primer synthesis by RNA polymerase', *Nature*, 439(7076), pp. 617–620. doi: 10.1038/nature04337.
- Zenkin, N., Severinov, K. and Yuzenkova, Y. (2015) 'Bacteriophage Xp10 anti-termination factor p7 induces forward translocation by host RNA polymerase.', *Nucleic acids research*. Oxford University Press, 43(13), pp. 6299–308. doi: 10.1093/nar/gkv586.
- Zenkin, N., Yuzenkova, Y. and Severinov, K. (2006) 'Transcript-Assisted Transcriptional Proofreading', *Science*. American Association for the Advancement of Science, 313(5786), pp. 518–520. doi: 10.1126/SCIENCE.1127422.
- Zhang, D. *et al.* (2016) 'Structural basis of prokaryotic NAD-RNA decapping by NudC', *Cell Research*, pp. 1062–1066. doi: 10.1038/cr.2016.98.
- Zhang, G. *et al.* (1999) 'Crystal Structure of *Thermus aquaticus* Core RNA Polymerase at 3.3 Å Resolution DNA template, and finally releases itself and the completed transcript from the DNA when a specific termination signal is encountered. The current view is that the', *Cell*, 98, pp. 811–824. doi: 10.1016/S0092-8674(00)81515-9.
- Zhang, H. *et al.* (2019) 'NAD tagSeq reveals that NAD⁺-capped RNAs are mostly produced from a large number of protein-coding genes in *Arabidopsis*.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 116(24), pp. 12072–12077. doi: 10.1073/pnas.1903683116.
- Zhang, J., Palangat, M. and Landick, R. (2010) 'Role of the RNA polymerase trigger loop in catalysis and pausing', *Nature Structural and Molecular Biology*, 17(1), pp. 99–105. doi: 10.1038/nsmb.1732.
- Zhang, N. and Buck, M. (2015) 'A Perspective on the Enhancer Dependent Bacterial RNA Polymerase', *Biomolecules*. Multidisciplinary Digital Publishing Institute (MDPI), 5(2), p. 1012. doi: 10.3390/BIOM5021012.
- Zhang, Y. *et al.* (2014) 'GE23077 binds to the RNA polymerase "i" and "i+1" sites and prevents the binding of initiating nucleotides.', *eLife*. eLife Sciences Publications, Ltd, 3, p. e02450. doi: 10.7554/eLife.02450.
- Zhao, J., Hyman, L. and Moore, C. (1999) 'Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis.', *Microbiology and molecular biology reviews : MMBR*. American Society for Microbiology, 63(2), pp. 405–45. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10357856> (Accessed: 12 October 2019).
- Zhilina, E. *et al.* (2012) 'Structural transitions in the transcription elongation complexes of bacterial RNA polymerase during σ -dependent pausing.', *Nucleic acids research*. Oxford University Press, 40(7), pp. 3078–91. doi: 10.1093/nar/gkr1158.
- Zhou, Y. *et al.* (2011) 'Determining the extremes of the cellular NAD(H) level by using an *Escherichia coli* NAD(+)-auxotrophic mutant.', *Applied and environmental microbiology*. American Society for Microbiology (ASM), 77(17), pp. 6133–40. doi: 10.1128/AEM.00630-11.
- Zumla, A. *et al.* (1999) 'The tuberculosis pandemic—Which way now?', *Journal of Infection*.

W.B. Saunders, 38(2), pp. 74–79. doi: 10.1016/S0163-4453(99)90072-5.

Zwerling, A. *et al.* (2011) ‘The BCG World Atlas: a database of global BCG vaccination policies and practices.’, *PLoS medicine*. Public Library of Science, 8(3), p. e1001012. doi: 10.1371/journal.pmed.1001012.

7 Bibliography

1. Julius, C. and Yuzenkova, Y. (2017) 'Bacterial RNA polymerase caps RNA with various cofactors and cell wall precursors', *Nucleic acids research*. Oxford University Press, 45(14), pp. 8282–8290. doi: 10.1093/nar/gkx452.
2. Julius, C., Riaz-Bradley, A. and Yuzenkova, Y. (2018) 'RNA capping by mitochondrial and multi-subunit RNA polymerases', *Transcription*. Taylor & Francis, 1264, pp. 1–13. doi: 10.1080/21541264.2018.1456258.
3. Julius, C. and Yuzenkova, Y. (2019) 'Noncanonical RNA-capping: Discovery, mechanism, and physiological role debate', *Wiley Interdisciplinary Reviews: RNA*. John Wiley & Sons, Ltd, 10(2), p. e1512. doi: 10.1002/wrna.1512.
4. Puiu, M and Julius C. (2019) 'Bacteriophage gene products as potential antimicrobials against tuberculosis.', *Biochemical Society transactions*. Portland Press Limited, 47(3), pp. 847–860. doi: 10.1042/BST20180506.
5. Manuscript submitted to Nucleic Acids Research: Manuscript ID NAR-03098-2019
Julius, C., Yuzenkova, Y. (2019) "Metabolic cofactors act as non-canonical initiating substrates for a primase and affect replication primer processing"

Bacterial RNA polymerase caps RNA with various cofactors and cell wall precursors

Christina Julius and Yulia Yuzenkova*

Centre for Bacterial Cell Biology, Institute for Cell and Molecular Biosciences, Newcastle University, Baddiley-Clark Building, Richardson Road, Newcastle upon Tyne, NE2 4AX, UK

Received January 20, 2017; Revised April 19, 2017; Editorial Decision May 07, 2017; Accepted May 09, 2017

ABSTRACT

Bacterial RNA polymerase is able to initiate transcription with adenosine-containing cofactor NAD⁺, which was proposed to result in a portion of cellular RNAs being ‘capped’ at the 5′ end with NAD⁺, reminiscent of eukaryotic cap. Here we show that, apart from NAD⁺, another adenosine-containing cofactor FAD and highly abundant uridine-containing cell wall precursors, UDP-Glucose and UDP-N-acetylglucosamine are efficiently used to initiate transcription *in vitro*. We show that the affinity to NAD⁺ and UDP-containing factors during initiation is much lower than their cellular concentrations, and that initiation with them stimulates promoter escape. Efficiency of initiation with NAD⁺, but not with UDP-containing factors, is affected by amino acids of the Rifampicin-binding pocket, suggesting altered RNA capping in Rifampicin-resistant strains. However, relative affinity to NAD⁺ does not depend on the −1 base of the template strand, as was suggested earlier. We show that incorporation of mature cell wall precursor, UDP-MurNAc-pentapeptide, is inhibited by region 3.2 of σ subunit, possibly preventing targeting of RNA to the membrane. Overall, our *in vitro* results propose a wide repertoire of potential bacterial RNA capping molecules, and provide mechanistic insights into their incorporation.

INTRODUCTION

For few decades multi-subunit RNA polymerase (RNAP) from *Escherichia coli* was known to be able to start RNA synthesis with cellular nucleotide coenzymes, adenosine derivatives NAD⁺ (nicotinamide adenine dinucleotide), NADH (reduced form of NAD⁺) and FAD (flavin adenine dinucleotide) (1). Authors of this work suggested that cofactor moiety could function analogously to the eukaryotic mRNA cap. Until very recently, the physiological significance of this discovery remained obscure, and the ac-

cepted view was that non-processed bacterial RNAs carry 5′ triphosphate. In the last few years however, the data started to accumulate that some cellular RNAs carry cap-like structure—*E. coli* and *Streptomyces venezuelae* bear NAD⁺ at the 5′ end (2). In 2015 those RNAs in *E. coli* were captured via 5′ NAD⁺ moiety and identified by next generation sequencing (3). It transpired that those RNAs were mainly regulatory sRNA and some mRNAs. Only relatively small proportion of the whole population of the particular RNA was NADylated *in vivo*. The most heavily NADylated were RNAI and CopA, abundant short antisense RNAs controlling pUC19 plasmid replication. Notably, for the NADylated transcripts with known start site, the +1 position coded for A, suggesting that it is RNAP incorporates NAD⁺ at the 5′ end of RNA via mechanism shown earlier, rather than some post-transcriptional mechanism being involved (2). While our paper was in preparation, biochemical and structural data using specific promoter assays has been published that confirmed the promoter-dependent and sequence-specific incorporation of NAD⁺, NADH and 3′-dephosphocoenzyme A (dpCoA) (4). Correlation between efficiencies of NAD⁺ incorporation *in vitro* and the extent of NADylation *in vivo* on two different promoters suggested that transcription might be the main, if not the only, capping mechanism. Crystal structures of initiation complex containing dinucleotide RNA products (to avoid confusion, here and after, we refer to the RNA length counting NAD⁺ and other dinucleotide co-factors as a single nucleotide) initiated with adenosine triphosphate (ATP), NAD⁺ and dpCoA were solved for *Thermus thermophilus* RNAP (4) demonstrated that, apart from interactions common for all three cofactors, contacts of NAD⁺ moiety additionally include side chains of β subunits residues D516 and H1237. The authors also proposed that nicotinamide moiety of NAD⁺ may rotate to interact with the −1 position of the template, thus explaining different efficiencies of NAD⁺ incorporation on different promoters.

Eukaryotic mRNA turnover depends on the efficiency of cap removal. Major catalytic role in decapping in eukaryotes is played by NUDIX motif-containing protein Dcp2p. In bacteria, NudC (NADH pyrophosphohydrolase), which contains NUDIX motif was shown to de-cap RNAs from

*To whom correspondence should be addressed. Tel: +44 191 208 3227; Fax: +44 191 208 3205; Email: y.yuzenkova@ncl.ac.uk

NAD⁺ (5). Notably, NudC has much higher affinity towards NADylated RNA compared to NAD⁺ itself (5). Existence of decapping mechanism makes the analogy between prokaryotic and eukaryotic RNA processing even stronger.

In eukaryotes mRNA capping plays vital role in RNA degradation, splicing, translation initiation and nuclear export. Physiological significance of bacterial RNA capping is not yet clear. The only role for capping that was put forward and got some experimental backing, is the protection of the transcript from degradation. The data on capped RNA stability are, however, conflicting. Bird *et al.* reported 3- to 4-fold increase of in NADylated RNAI stability in Δ NudC cells (lacking de-capping activity) (4), in contrast to another study, where deletion of NudC did not affect the overall stability of RNAI and GcvB, two RNAs, most heavily NADylated *in vivo* (3). Moreover, some RNAs with high NAD⁺ cap content are inherently very stable, for example *sroB* with half-life of more than 32 minutes (6). All these data suggest possible additional roles for prokaryotic capping apart from RNA stability.

To date only ADP analogues were identified as caps, NAD⁺ and/or NADH. Cells use a number of nucleotide cofactors and these might be just the first identified ones among many substrates used by RNAP for RNA capping. There are several poorly characterized NUDIX hydrolases in *E. coli*, which can potentially serve to remove capping molecules different from NAD⁺ (7), which have not been found yet.

Here we provide further insights into the mechanism of capping with adenine containing cofactors (NAD⁺, NADH, NADP⁺, FAD) and its possible role in transcription. We also show that the repertoire of potential capping molecules is wider, and includes uridine containing precursors of cell wall synthesis (UDP-glucose and UDP-N-acetylglucosamine). Our data also suggest the role of region 3.2 of initiation factor sigma in guarding transcription against incorporation of 'long-tailed' NTP analogues into RNA.

MATERIALS AND METHODS

Materials

ATP and UTP were from GE Healthcare; AMP, ADP, NAD⁺, NADH, NADP⁺, FAD, UMP, UDP, UDP-Glc and UDP-GlcNAc were from Sigma Aldrich; UDP-MurNAc pentapeptide was a kind gift from Prof. Vollmer.

Proteins

Mutations in *E. coli* *rpoB* gene were constructed by site-directed mutagenesis in polycistronic expression plasmid pGEMABC, coding for RNAP core subunits α , β and β' (8). Those subunits were overexpressed in *E. coli* T7 express strain (New England Biolabs) together with ω subunits from expression plasmid pRSFD_2_*rpoZ* (8). Wild-type and mutant RNAPs core enzymes were purified as described in (9). RpoS gene encoding σ^S was cloned into expression vector pET28, as previously were σ^{70} and $\sigma^{70\Delta 3.2}$ (10). N-terminal Hisx6-tagged σ^{70} , $\sigma^{70\Delta 3.2}$, σ^S were ex-

pressed from overexpression vector pET28 and purified as described in (11).

In vitro transcription

A total of 0.3 pmols of wild-type or mutant RNAP core with 1 pmols of σ^{70} (wild-type or mutant) or σ^S and 2 pmols of promoter-containing linear DNA fragment were incubated at 37°C for 10 min in 10 μ l of transcription buffer (20 mM Tris-HCl (pH 7.9), 40 mM KCl, 0.1 mM ethylenediaminetetraacetic acid) at 37°C, then nucleotides or nucleotides analogues were added to the final concentration of 500 μ M (unless otherwise indicated). Transcription was initiated by the addition of 10 mM MgCl₂, 50 μ M (α^{32} P)-CTP, 12.5 Ci/mmol (Hartmann Analytic) for RNAI and *acnA* templates; for T7A1 template 50 μ M (α^{32} P)-UTP, 12.5 Ci/mmol of were used. Reactions were stopped after 3 min incubation at 37°C (unless otherwise indicated) by the addition of formamide-containing loading buffer. For the kinetics of synthesis of 9 nt-long transcript from RNAI_{mod}, 500 μ M ATP or NAD⁺ were incubated with promoter complex and then 20 μ M UTP and 20 μ M (α^{32} P)-CTP, 12.5 Ci/mmol were added in the absence or presence of 5 μ g/ml of Rifampicin. Reactions were stopped after periods of time indicated on Figure 4A. Products were separated on denaturing polyacrylamide gels (30% acrylamide, 3% bis-acrylamide, 6M urea, 0.5 Tris-borate EDTA buffer), revealed by PhosphorImaging (GE Healthcare), and analysed using ImageQuant software (GE Healthcare). For 9-nt long RNA synthesis in Figure 2D on RNAI_{mod} template 500 μ M of either ATP, NAD⁺, NADH, NADP or FAD were incubated with promoter complex and then 25 μ M UTP and 25 μ M (α^{32} P)-CTP, 12.5 Ci/mmol, were added and reactions were incubated for 10 min. For similar experiment on Figure 5C to synthesize 12-nt long RNA on *acnA* template 500 μ M of either UTP, UDP-Glc or UDP-GlcNAc were incubated with promoter complex and then 25 μ M ATP and 25 μ M (α^{32} P)-CTP, 12.5 Ci/mmol, were added and reactions were incubated for 10 min. For apparent K_m determination on RNAI_{mod} and *acnA* templates NTPs and analogues were used in concentrations ranging from 10 μ M to 10 mM and constant 50 μ M CTP (second NTP for both templates) concentration. Reactions were incubated for periods of time chosen to get approximately the same intensities of the product bands. The bands intensities were quantified using ImageQuant software; to calculate the initial reaction rate these numbers were divided by reaction duration time. These data were fitted to hyperbolic equation using non-linear regression procedure in SigmaPlot software.

RESULTS

Escherichia coli RNAP is able to initiate transcription using adenosine diphosphate analogues

Abortive production of short oligonucleotides is the very first stage of the synthesis of the RNA. These products can be resolved in high density denaturing polyacrylamide gel. We analysed the efficiency of incorporation of potential cap molecules into RNA in abortive synthesis by *E. coli* RNAP σ^{70} holoenzyme on a linear DNA containing

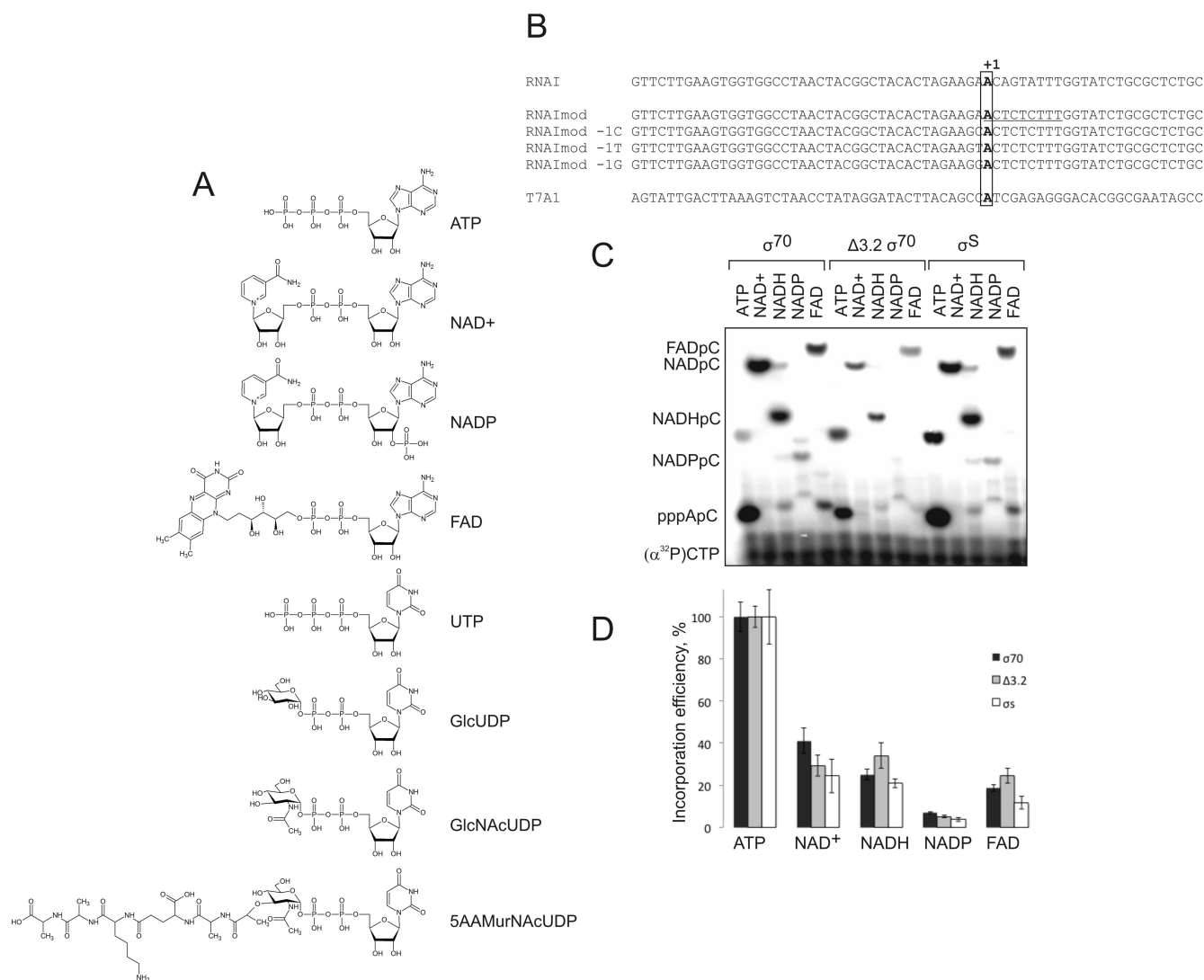


Figure 1. ADP-related cellular cofactors are utilized by RNAP for initiation of transcription. (A) Structures of ATP, NAD⁺, NADP⁺, FAD, UDP, UDP-Glc (GlcUDP), UDP-GlcNAc (GlcNAcUDP) and UDP-MurNAc pentapeptide (5AAMurNAcUDP). (B) Templates (partial sequence of non-template strand around transcription start site) used for *in vitro* experiments containing T7A1, RNAI promoters and RNAI template with modified initially transcribed sequence, RNAImod. +1 position for all templates is in bold, -1 position is framed, 9 nt initially transcribed sequence in RNAImod is underlined. (C) Initial transcript synthesis on RNAImod template using RNAP holo σ^{70} , holo σ^S and holo $\sigma^{70\Delta 3.2}$ and 500 μ M ATP, NAD⁺, NADH, NADP⁺ and FAD as initiating substrates and (α^{32} P)-CTP as the next nucleotide. (D) Plot reflecting incorporation efficiencies for alternative substrates in percentage from efficiency of ATP incorporation for RNAP holo σ^{70} , holo σ^S and holo $\sigma^{70\Delta 3.2}$ on RNAImod, the values are an average of the three independent experiments, error bars represent plus and minus one standard deviation.

promoter RNAI (that was shown to be the most heavily modified by NAD⁺ in *E. coli*). ATP or co-enzymes (each at 500 μ M concentration) (structures on Figure 1A) were extended by (α^{32} P)-CMP (at 50 μ M). Position +3 of the RNAI promoter was mutated to T to make RNAImod template (here and after we refer to the sequence of the non-template strand) to preclude extension of RNA further than position +2, which would obscure kinetics analysis (Figure 1B). As can be seen from Figure 1C, NAD⁺, NADH, FAD can be efficiently incorporated into the 2-nt long transcript. The results are consistent with the recently published work (4) and the earlier study (1). Efficient incorporation of these coenzymes into the transcript is not a particular property of RNAI, as they were similarly efficiently extended by (α^{32} P)-

UMP on a strong promoter A1 from bacteriophage T7, where transcription also starts with A (Figure 2A). NADP⁺ was incorporated very inefficiently into the transcript on both templates, likely due to the 2'-phosphate group sterical hindrance during phosphodiester bond formation. At 1 mM initiating NAD⁺ and ATP on RNAImod template, the kinetics of formation of NAD⁺pC was comparable (~2 times slower) to that of pppApC (Figure 2B). The K_m s for ATP, NAD⁺ and NADH in initiation on RNAImod template were $90 \pm 11 \mu$ M, $358 \pm 67 \mu$ M and $380 \pm 72 \mu$ M, respectively, which is much lower than their cellular concentrations (see 'Discussion' section).

In the presence of CTP and UTP, the short transcripts initiated with NAD⁺, NADH and FAD on the RNAImod

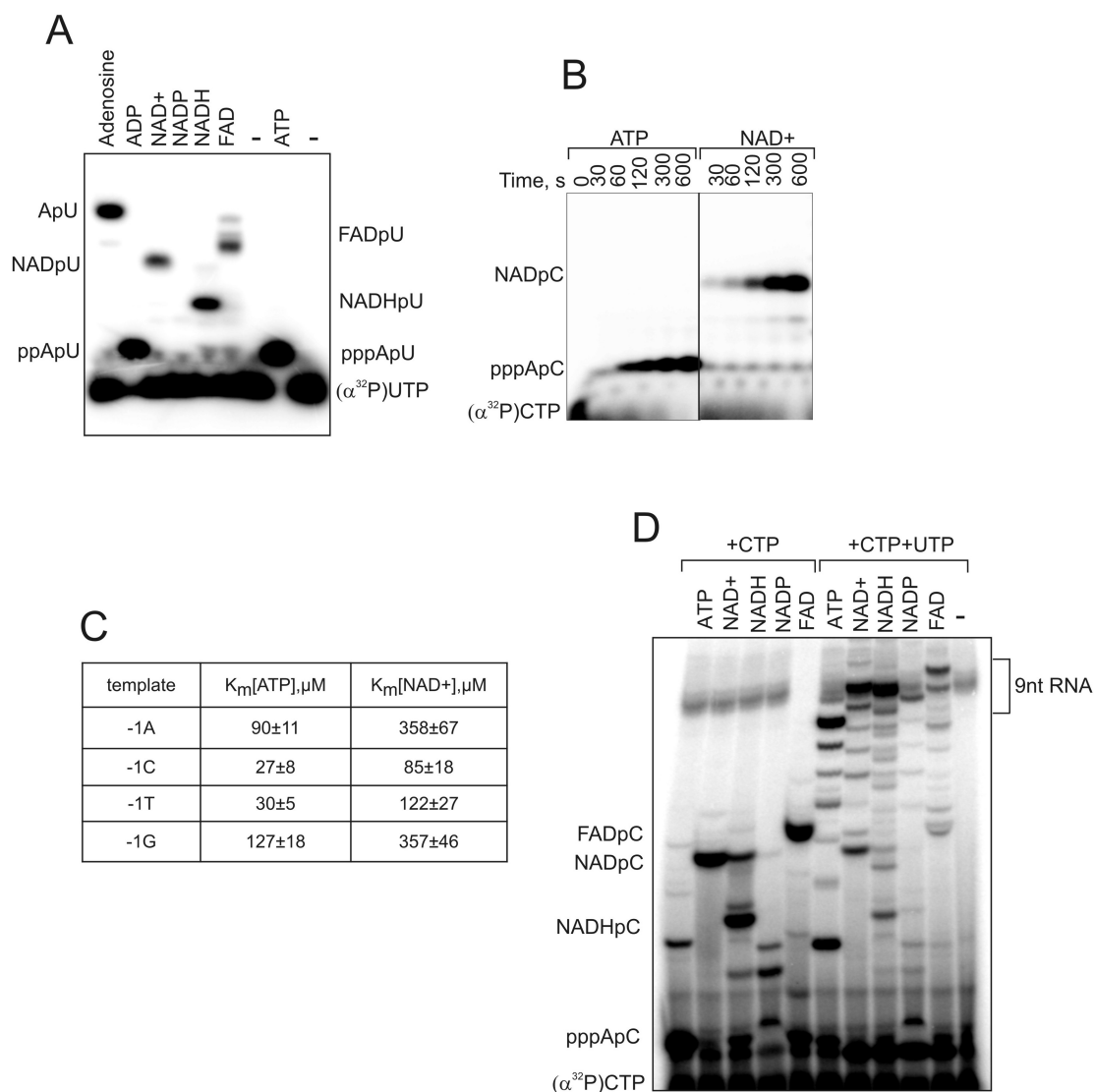


Figure 2. Biochemical characterization of the ADP analogues incorporation into a transcript. (A) Initial products synthesized using ATP and cofactors as initiating substrates on T7A1 promoter template with RNAP holo^{*σ*70} *Escherichia coli* and (α^{32} P)-UTP as next nucleotide. (B) Kinetics of initial product synthesis on RNAImod template with RNAP holo^{*σ*70} with either ATP or NAD⁺ as initiating substrate. (C) K_m for ATP and NAD⁺ as initiating substrates measured on RNAImod templates with different identity of −1 template bases (sequences in Figure 1B); numbers that follow the \pm sign are errors that are standard deviations of the fitting. (D) Transcripts initiated with ATP, NAD⁺, NADH, NADP⁺ and FAD, are elongated to 9 nt transcript on RNAImod template.

template were efficiently extended into 9-nt products (Figure 2D). The amounts of 9-nt products were comparable to those initiated with ATP, suggesting that, similar to NAD⁺ capping, NADH and FAD capping may exist in the cell. Note also that the efficiency of abortive products extension (the ratio between 9-nt RNA and dinucleotides) in the case of co-factors was higher than in the case of ATP, suggesting that co-factors increase efficiency of promoter escape (see below).

The −1 position of template does not influence the relative efficiency of NAD⁺ versus ATP incorporation into the transcript

Nicotinamide moiety may, in theory, form base pair with

template DNA base at the position −1. This may change specificity of NADylation at promoters with particular base at −1 position. It was observed that A to C (non-template strand) change in −1 position decreased the overall efficiency of NADylated transcript synthesis (4). To test if the base at position −1 affects the affinity to NAD⁺, we used linear templates with RNAImod promoter variants with changes in the −1 position (Figure 1B). We measured apparent K_m for NAD⁺ and ATP on these four promoters variants in abortive synthesis as above (Figure 2C). We found that K_m for NAD⁺ incorporation in transcript was lower for templates with C and T in −1 position compared with values for −1 A or G. However, the same tendency was observed for incorporation of ATP, suggesting that position −1 of template does not specifically affect affinity to

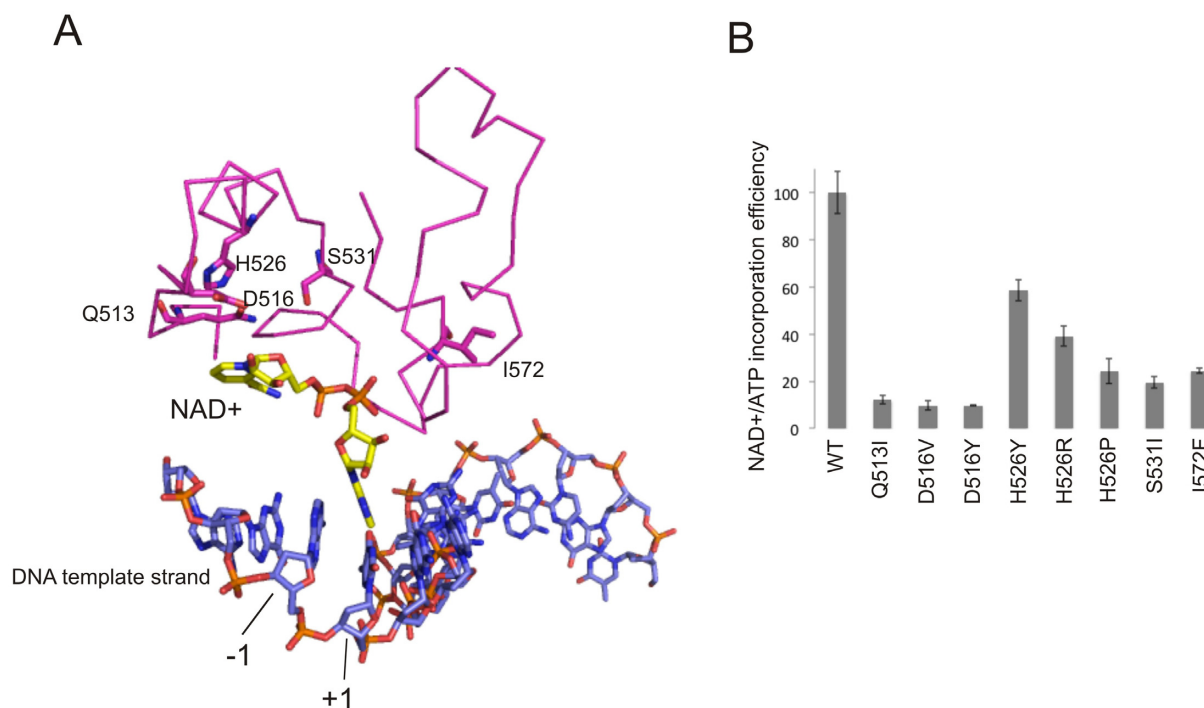


Figure 3. Aminoacids of Rifampicin-binding pocket of β subunit of RNAP influence incorporation of ADP analogues into a transcript. (A) Crystal structure of the initiation complex with NAD⁺ pC product with *Thermus thermophilus* RNAP, adapted from PDB ID: 5D4D (4). The amino acids of Rif-pocket, whose changes were tested *in vitro* are in purple (*Escherichia coli* numbering), NAD⁺ is in yellow, DNA is in blue, -1 and +1 positions of the templates are indicated. (B) Relative efficiency of incorporation of 500 μ M NAD⁺ versus ATP into dinucleotide product on RNAImod template by WT and mutant RNAPs with aminoacid changes in β subunit indicated below the plot. The values are an average of the three independent experiments, error bars represent plus and minus one standard deviation.

NAD⁺. This result is consistent with crystal structures of the initiating complexes where initiating NAD⁺ makes the same contacts with template DNA as initiating ATP (4). Therefore it is likely that different steps of initiation, rather than direct interaction with the template, may influence the relative efficiency of initiation with NAD⁺ on templates with different -1 position (4). Note that the measured K_m s for NAD⁺ on either of the templates are much lower than the cellular concentrations (see 'Discussion' section).

Core RNAP determinants for NAD⁺ incorporation

We tested if there are any other structural determinants for cofactors incorporation apart from -1 position of template. 5' NTP of the growing transcript passes through the Rifampicin-binding pocket of the β subunit of RNAP. It was shown that γ -phosphate of the initiating ATP is in proximity to the β subunit fragment between amino acids 516 and 540 (12), suggesting where the nicotinamide group of NAD⁺ could be bound during initiation. We tested several RNAPs with amino acid changes in Rifampicin-binding pocket (Rif pocket) — Q513L, F514A, D516V, D516Y, H526Y, H526R, H526P, S531L, N568A, I572F on RNAImod promoter with either NAD⁺ or ATP as an initiating substrate. The activities of the enzymes in abortive synthesis varied significantly (either due to specific activities of the enzymes or distortion of the Rif pocket). Figure 3B shows relative efficiencies of NAD⁺ incorporation (NAD⁺/ATP ratio) by the mutant enzymes in percentage

from that of the WT RNAP. Indeed, in agreement with our hypothesis that part of NAD⁺ maybe bound in the Rif pocket, the effect was dependent both on the position and nature of the amino acid substitution. The strongest effect on NAD⁺ utilization was produced by mutations of D516 of β (Figure 3B). This corroborates the crystal structure of the initiating complex with NAD⁺ primed RNA product, where the D516 side chain is in close proximity of the nicotinamide moiety of the NAD⁺ (4) (Figure 3A). Changes of other aminoacids that are too far to interact with NAD⁺ directly, perhaps affect the overall shape of the Rif-pocket.

Initiation factor does not play a role in selectivity of ADP-containing co-enzymes as substrates

A larger proportion of RNAI was found to be NADylated in stationary growth phase compared to the exponential phase (3). In stationary phase most transcripts are made by RNAP holoenzyme containing σ^S , while housekeeping σ^{70} operates in exponential phase. We analysed if different initiation factors may dictate specificity towards capping cofactors. To answer this question we tested initiation with NAD⁺ on RNAI promoter with holo σ^{70} and holo σ^S (Figure 1C and D). The rates of abortive products formation with ATP, NAD⁺, NADH, NADP⁺ and FAD were similar for both RNAPs (Figure 1C and D). K_m values for NAD⁺ were also close for both holoenzymes ($358 \pm 67 \mu$ M for holo σ^{70} and $352 \pm 88 \mu$ M for holo σ^S). The results suggest that at least the two initiation factors tested do not provide signif-

icant specificity towards utilization of NAD⁺ as initiation substrate.

Previous functional and structural analyses suggested that region 3.2 of σ subunit protrudes towards catalytic site of the RNAP and may contact the 5' end of short transcripts (10,11,13,14). We tested initiation with NAD⁺, NADH, NADP⁺ and FAD (500 μ M) on RNAI promoter by holoenzymes containing either wild-type σ^{70} or mutant σ^{70} lacking region 3.2, $\sigma^{70\Delta 3.2}$ (Figure 1C and D). We did not observe any significant differences in specificity, suggesting that region 3.2 does not make contacts critical for cofactors' binding.

NADylation of transcript stimulates escape of the RNAP from promoter

From our results it follows that NAD⁺ interacts differently with RNAP as compared to ATP. These differences might affect stability of the short abortive transcripts, and as a result, their extension during promoter escape. We therefore tested kinetics of 9-nt RNA production initiated with either ATP or NAD⁺ (500 μ M) on RNAImod template. A relatively low concentration of CTP and UTP (20 μ M) allowed us to monitor the accumulation of short RNAs ranging from 2 to 9 nt in length. As can be seen from Figure 4A and B there were much less accumulation of the 2- and 3-nt long transcripts when transcription was initiated with NAD⁺ as compared to ATP (compare the traces of 600 s products in Figure 3B). Rifampicin is known to block escape into elongation, with concomitant increase of abortive synthesis (15). We hypothesized that, if NAD⁺ containing initial transcripts are bound tighter by RNAP than ATP containing ones, Rifampicin may have different effect on their release. We used low concentration of Rifampicin (5 μ g/ml), which was not enough to block transcription completely. Addition of Rifampicin to reaction inhibited production of 9-nt RNA initiated by either ATP or NAD⁺. Also, as expected Rifampicin increased release of pppApC dinucleotide. However, Rifampicin almost didn't affect production of NAD⁺ containing short RNAs, highlighting the difference between promoter escape with ATP and NAD⁺ (compare the traces of 600 s products in Figure 3B). These results suggest that NAD⁺ moiety at the 5' end indeed can stabilize short transcripts in the RNAP active centre.

UDP derivatives can serve as initiating substrates for transcription

Exponentially growing in rich medium *E. coli* cells contain high concentrations of a number of nucleotide analogues potentially capable of initiating transcription. One of the most abundant small molecules in *E. coli* cell, second only to ATP, is UDP-N-acetylglucosamine (UDP-GlcNAc) (16), which participates in formation of peptidoglycan, lipopolysaccharides and teichoic acid cell wall components. We tested if UDP-GlcNAc, along with another precursors of cell wall synthesis, Uridine 5'-diphosphoglucose (UDP-Glc) (structures in Figure 1A) can initiate transcription. We also tested a more complex compound, UDP-MurNAc pentapeptide (Figure 1A), the last-step precursor before the formation of Lipid I, a building block of the cell wall.

In this experiment we used linear templates containing *acnA* promoter, on which transcription starts with UTP (Figure 5A). We analyzed initiation from UDP-Glc, UDP-GlcNAc, UDP-MurNAc pentapeptide (5AA-MurNAc), along with UMP, UDP and UTP as controls (each at 500 μ M concentrations) (Figure 5A) by monitoring synthesis of the dinucleotide transcript after addition of an (α^{32} P)-CMP (at 50 μ M concentration). As can be seen from Figure 5A, *E. coli* RNAP can efficiently incorporate UDP-Glc and UDP-GlcNAc (comparably to UTP; lanes 3–5). The K_m values on the *acnA* promoter for UDP-Glc and UDP-GlcNAc were 300 ± 62 μ M and 333 ± 41 μ M, respectively (compared to 120 ± 17 μ M for UTP). These K_m values are well below intracellular concentrations of UDP-Glc and UDP-GlcNAc in exponential growth phase (2.5 and 9.2 mM, respectively). We analysed if dinucleotides initiated with UDP-Glc and UDP-GlcNAc can be extended to facilitate escape into elongation. We mutated *acnA* promoter (*acnAmod*; Figure 5A) so that we could monitor extension to 12-nt RNA in the presence of the ATP and CTP. As can be seen from Figure 5C, the transcripts initiated with UDP-Glc and UDP-GlcNAc were elongated by RNAP to the 12 nt. Note also, that the efficiency of promoter escape with UDP-Glc and UDP-GlcNAc was higher than that with UTP. These findings suggest that the cell wall precursors UDP-Glc and UDP-GlcNAc may serve as an RNA caps *in vivo*.

Interestingly, UDP-MurNAc pentapeptide was not utilized by RNAP at all. It is possible that RNAP possesses a mechanism that prevents incorporation of UDP-MurNAc pentapeptide into an RNA, which would be too costly for cells and may lead to unwanted targeting of the modified transcript towards the membranes. One of the possible obstacles for the pentapeptide in the active site could be region 3.2 of σ^{70} . We therefore analysed incorporation of UMP, UDP, UTP, UDP-Glc, UDP-GlcNAc and UDP-MurNAc pentapeptide (500 μ M) by holoenzymes formed with wild-type σ^{70} or $\sigma^{70\Delta 3.2}$. As can be seen from Figure 5A, the mutant RNAP indeed acquired partial ability to incorporate UDP-MurNAc pentapeptide, while usage of smaller cell wall precursors was not affected (see also relative efficiencies of dinucleotide formation in Figure 5B). This result suggests that region 3.2 may participate in guarding against incorporation of the cellular nucleotide analogues with long side chains.

To test if Rif-pocket plays role in selectivity of UDP-GlcNAc utilization in initiation, as it does for ADP-containing co-factors, we tested the activities of Rifampicin resistant mutant RNAPs that we used above. Mutant RNAP with Q513I substitution was inactive on *acnA* even with UTP, and was excluded from analysis. The rest of mutations in Rif-pocket did not affect incorporation of UDP-GlcNAc into the transcript strongly (Figure 5D), suggesting that UDP-GlcNAc may not make specific contacts with amino acids of the Rif-pocket.

DISCUSSION

Here we showed that RNAP is able to incorporate variety of cellular cofactors at the 5' end of the transcript, suggesting a possibility of a wide repertoire of RNA caps in bacteria. In addition to NAD⁺ that had been shown to cap some

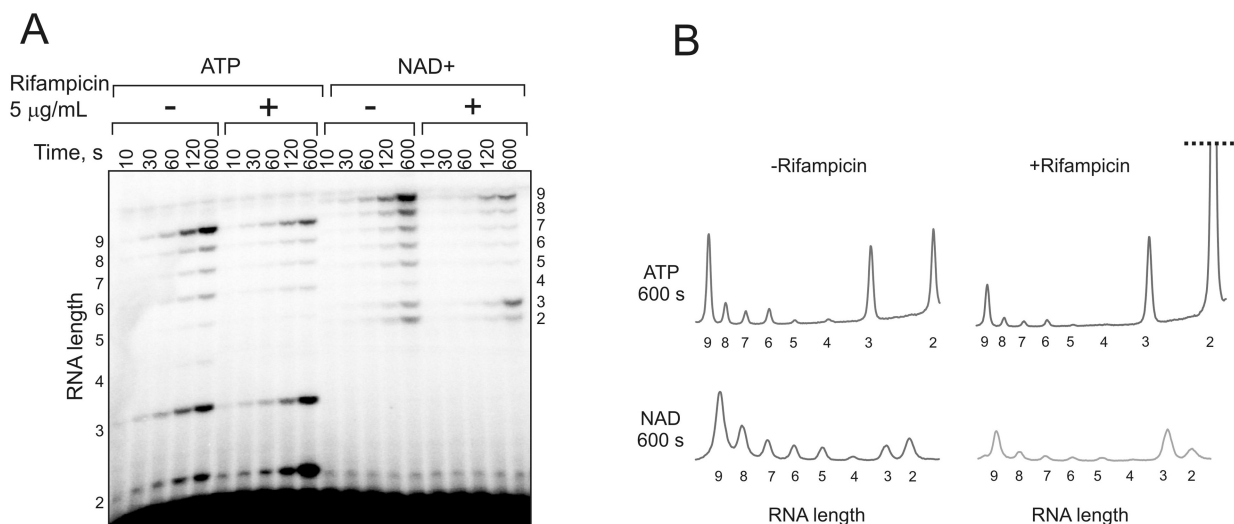


Figure 4. NAD⁺ as initiating substrate improves escape of RNAP from promoter compared to ATP. (A) Kinetics of synthesis of 9 nt transcript on RNAI mod promoter template (Figure 1B) using either ATP or NAD⁺ at 1 mM, 20 μ M (α^{32} P)-CTP and 20 μ M UTP in the absence or presence of 5 μ g/ml of Rifampicin. (B) Signal traces taken across the 600 s bands.

of bacterial RNAs *in vivo*, another adenine containing co-factor FAD can also be incorporated by RNAP in sequence-dependent manner, and thus cap RNAs on +1A (T in the +1 position of the template strand) promoters. Furthermore, we show that UDP-containing cell wall precursors may efficiently cap RNAs on +1U promoters.

The crystal structure of promoter complex with NADpC dinucleotide transcript showed the interactions of nicotinamide moiety with the Rif-pocket of β subunit (4). Indeed, all amino acid substitutions in the Rif-pocket we tested in our experiments specifically decreased the ability of RNAP to incorporate NAD⁺, thus confirming the structural observations. We observed that substitutions at position 516 are the most detrimental for NAD⁺ incorporation, consistent with the contacts of nicotinamide moiety with β D516 seen in the structure. Interestingly, Rifampicin resistant substitutions at positions 516, 526 and 531 that strongly inhibited NAD⁺ incorporation, are those most frequent found in clinic (17). The control of the RNA capping by the Rif-pocket may be involved also after the cap binding in the active centre, at the stage of RNA extension and promoter escape. Indeed, we found that transcription started with NAD⁺ produced far less of short abortive products than transcription started with ATP on the same promoter.

The crystal structure of the promoter complex with the NADpC revealed that nicotinamide moiety does not interact with the template (4). The authors however observed that a change in the -1 position of the template may change the proportion of NADylated RNAs when both ATP and NAD⁺ are used in the reaction. They suggested that nicotinamide moiety might change conformation to interact with the template base in -1 position, which has not been captured in the crystals. Although we found that K_m for NAD⁺ for promoters with -1C or -1T were lower compared to the promoters with -1G or A, the same tendency was observed for ATP, suggesting that the preference of NAD⁺ for -1 position is not explained by specific base-pairing of nicoti-

namide moiety with the -1 position of the template. We therefore suggest that -1 position of the promoter affects the properties of other step(s) of initiation, thus affecting the efficiency of RNA NADylation.

Region 3.2 of the σ subunit is in close proximity to the Rif-pocket (8) and may serve as another determinant for an efficiency of incorporation of the cofactors. However, we did not find any significant effect of the deletion of region 3.2 on incorporation of either NAD⁺, NADH, FAD or NADP, as compared to ATP. We also observed no difference in efficiency of usage of these factors by RNAP equipped with either housekeeping σ^{70} versus stationary phase σ^S initiation factor. These results suggest that σ subunits do not directly interact with the initiating co-enzymes. We however cannot exclude that other alternative initiation factors can contribute towards or against incorporation of non-canonical capping nucleotides analogues. Furthermore, σ region 3.2 may affect the relative efficiency of capping by its interactions with the template DNA upstream of the start site, thus affecting escape into productive elongation.

We further extended the potential bacterial RNA cap repertoire to derivatives of UDP, the highly abundant precursors of cell wall components. Interestingly we found that while σ region 3.2 did not have effect on the incorporation of UDP-Glc and UDP-GlcNAc, it did inhibit incorporation of a larger UDP-MurNAc pentapeptide. The protection of more complex cell wall precursor from incorporation into RNA may have biological significance. The more complex precursors are more expensive for the cell than initial precursors UDP-Glc and UDP-GlcNAc. Furthermore, modification of the 5' end of RNA with UDP-MurNAc pentapeptide may potentially lead to targeting of an RNA to a membrane, thus affecting its expression.

K_m values for NAD⁺ are in the range of 100–400 μ M on various promoter variants, which is below published intracellular concentration of 2.3 mM in exponential *E. coli* cells

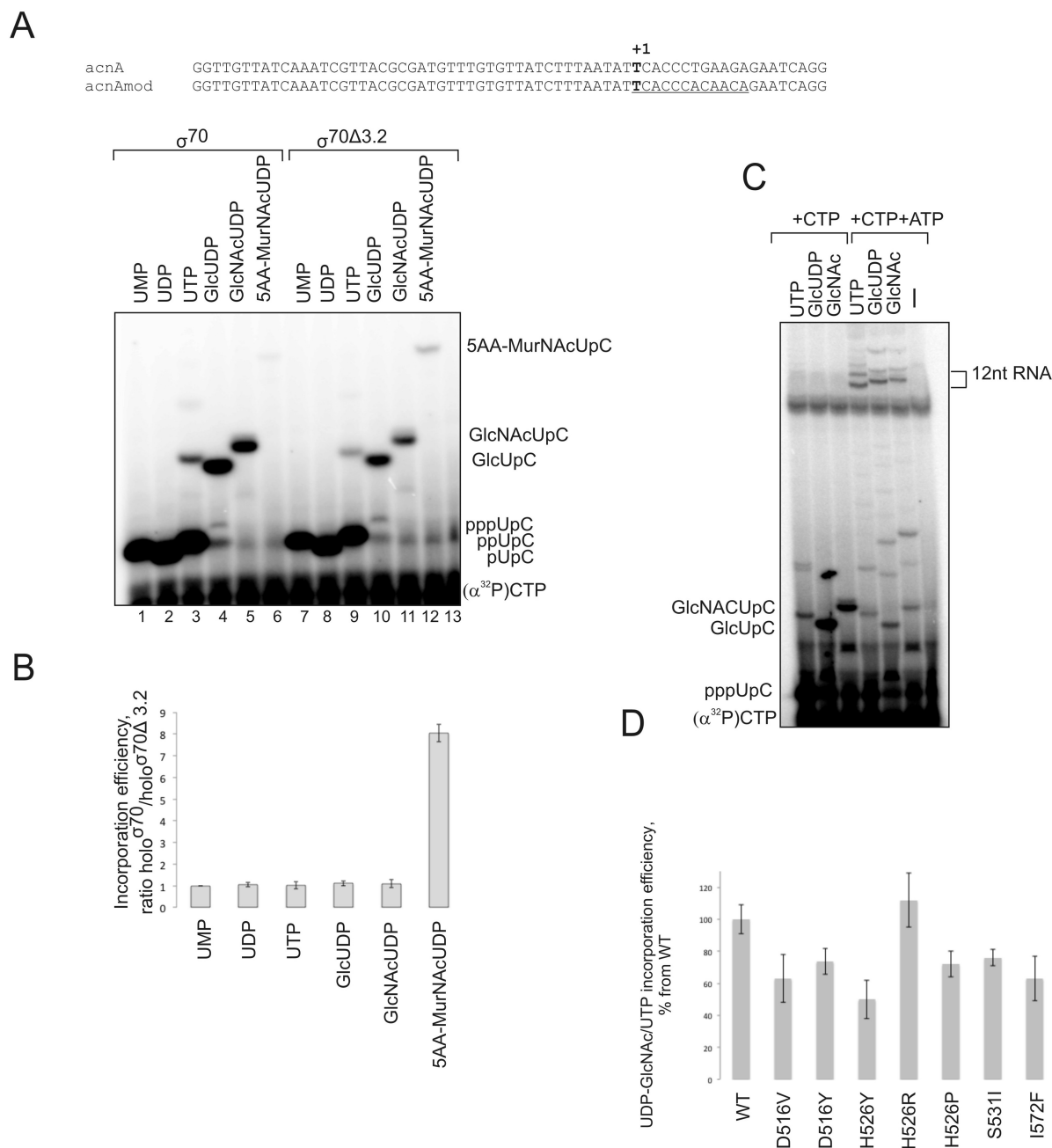


Figure 5. Analogues of UDP can serve as initiating substrates for RNAP *in vitro*. (A) Template with acnA promoter (partial sequence around transcription start site) and acnAmod template, +1 position is in bold, 12 nt initially transcribed sequence in RNAImod is underlined. Dinucleotide synthesis on acnA template using either RNAP holo σ^{70} or holo $\sigma^{70\Delta 3.2}$, and 500 μ M UMP, UDP, UTP, UDP-Glc, UDP-GlcNAc and UDP-MurNAc pentapeptide as initiating substrates and (α^{32} P)-CTP as the next nucleotide. (B) Plot below the gel represents the ratio of incorporation of indicated substrates by holo $\sigma^{70\Delta 3.2}$ versus holo σ^{70} , the values are an average of the two independent experiments, error bars represent plus and minus one standard deviation. (C) Transcripts initiated with UTP, UDP-Glc, UDP-GlcNAc, are elongated to 12 nt transcript on RNAImod template by WT and mutant RNAPs with aminoacid changes in Rifampicin-binding pocket of β subunit indicated below the plot. (D) Relative efficiency of incorporation of 500 μ M UDP-GlcNAc in comparison to UTP into dinucleotide product on RNAImod template by WT and mutant RNAPs with aminoacid changes in Rifampicin-binding pocket of β subunit indicated below the plot. The values are an average of the two independent experiments, error bars represent plus and minus one standard deviation.

grown with glucose (16). K_m values for the UDP-GlcNAc and UDP-Glc (both in the range of $\sim 300 \mu\text{M}$) on *acnA* promoter are also far lower the cellular concentration of these compounds (~ 9 and 2.3 mM , respectively) (16), and thus are even more favorable for RNA capping. In addition to NudC, which was shown to uncap NAD-RNAs, there are 13 poorly characterized NUDIX hydrolases in *E. coli*, which can potentially remove capping molecules, hypothetically including FAD and derivatives of UDP (7).

There might be an unexpected connection between capping (and thus potentially gene expression) and the effect of various antibiotics. For example, inhibition of protein synthesis by chloramphenicol or tetracycline, as well as inhibition of cell wall biosynthesis by fosfomycin lead to significant increase of the UDP-GlcNAc concentration (18), potentially changing capping efficiency for transcripts produced from +1U promoters. All tested Rifampicin resistant mutants of RNAP have impaired NAD⁺ capping activity most likely due to an altered geometry of Rif-pocket. Therefore the reduced fitness of Rifampicin resistant mutants (19), in part, may be attributed to the altered NADylation of RNA resulting in aberrant regulation of RNA stability and thus gene expression. Furthermore, subinhibitory concentrations of Rifampicin may have different effects on production of RNA initiated with NTP and a co-factor, as can be seen from Figure 4A.

Overall our work provides mechanistic insight into the process of prokaryotic capping of RNA by RNAP with the variety of cellular cofactors *in vitro*.

ACKNOWLEDGEMENTS

The authors thank Prof. Waldemar Vollmer for UDP-MurNAc pentapeptide.

FUNDING



Newcastle University Faculty of Medical Sciences Studentship (to C.J.); Royal Society University Research Fellowship (to Y.Y.). Funding for open access charge: the UK Biotechnology and Biological Sciences Research Council.

Conflict of interest statement. None declared.

REFERENCES

- Malygin, A.G. and Shemyakin, M.F. (1979) Adenosine, NAD and FAD can initiate template-dependent RNA synthesis catalyzed by *Escherichia coli* RNA polymerase. *FEBS Lett.*, **102**, 51–54.
- Chen, Y.G., Kowtoniuk, W.E., Agarwal, I., Shen, Y. and Liu, D.R. (2009) LC/MS analysis of cellular RNA reveals NAD-linked RNA. *Nat. Chem. Biol.*, **5**, 879–881.
- Cahova, H., Winz, M.L., Hofer, K., Nubel, G. and Jaschke, A. (2015) NAD captureSeq indicates NAD as a bacterial cap for a subset of regulatory RNAs. *Nature*, **519**, 374–377.
- Bird, J.G., Zhang, Y., Tian, Y., Panova, N., Barvik, I., Greene, L., Liu, M., Buckley, B., Krasny, L. and Lee, J.K. (2016) The mechanism of RNA 5' capping with NAD⁺, NADH and desphospho-CoA. *Nature*, **535**, 444–447.
- Hofer, K., Li, S., Abele, F., Frindert, J., Schlotthauer, J., Grawenhoff, J., Du, J., Patel, D.J. and Jaschke, A. (2016) Structure and function of the bacterial decapping enzyme NudC. *Nat. Chem. Biol.*, **12**, 730–734.
- Vogel, J., Bartels, V., Tang, T.H., Churakov, G., Slagter-Jager, J.G., Huttenhofer, A. and Wagner, E.G. (2003) RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res.*, **31**, 6435–6443.
- Jaschke, A., Hofer, K., Nubel, G. and Frindert, J. (2016) Cap-like structures in bacterial RNA and epitranscriptomic modification. *Curr. Opin. Microbiol.*, **30**, 44–49.
- Murakami, K.S. (2013) X-ray crystal structure of *Escherichia coli* RNA polymerase sigma70 holoenzyme. *J. Biol. Chem.*, **288**, 9126–9134.
- Orlova, M., Newlands, J., Das, A., Goldfarb, A. and Borukhov, S. (1995) Intrinsic transcript cleavage activity of RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 4596–4600.
- Zenkin, N. and Severinov, K. (2004) The role of RNA polymerase sigma subunit in promoter-independent initiation of transcription. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 4396–4400.
- Kulbachinskiy, A. and Mustaev, A. (2006) Region 3.2 of the sigma subunit contributes to the binding of the 3'-initiating nucleotide in the RNA polymerase active center and facilitates promoter clearance during initiation. *J. Biol. Chem.*, **281**, 18273–18276.
- Severinov, K., Mustaev, A., Severinova, E., Kozlov, M., Darst, S.A. and Goldfarb, A. (1995) The beta subunit Rif-cluster I is only angstroms away from the active center of *Escherichia coli* RNA polymerase. *J. Biol. Chem.*, **270**, 29428–29432.
- Severinov, K., Fenyo, D., Severinova, E., Mustaev, A., Chait, B.T., Goldfarb, A. and Darst, S.A. (1994) The sigma subunit conserved region 3 is part of "5'-face" of active center of *Escherichia coli* RNA polymerase. *J. Biol. Chem.*, **269**, 20826–20828.
- Vassilyev, D.G., Sekine, S., Laptchenko, O., Lee, J., Vassilyeva, M.N., Borukhov, S. and Yokoyama, S. (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature*, **417**, 712–719.
- Campbell, E.A., Korzhcheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A. and Darst, S.A. (2001) Structural mechanism for rifampicin inhibition of bacterial rna polymerase. *Cell*, **104**, 901–912.
- Bennett, B.D., Kimball, E.H., Gao, M., Osterhout, R., Van Dien, S.J. and Rabinowitz, J.D. (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.*, **5**, 593–599.
- Cavusoglu, C., Hilmioglu, S., Guneri, S. and Bilgic, A. (2002) Characterization of *rpoB* mutations in rifampin-resistant clinical isolates of *Mycobacterium tuberculosis* from Turkey by DNA sequencing and line probe assay. *J. Clin. Microbiol.*, **40**, 4435–4438.
- Mengin-Lecreulx, D. and van Heijenoort, J. (1990) Correlation between the effects of fosfomycin and chloramphenicol on *Escherichia coli*. *FEMS Microbiol. Lett.*, **54**, 129–133.
- Melnik, A.H., Wong, A. and Kassen, R. (2015) The fitness costs of antibiotic resistance mutations. *Evol. Appl.*, **8**, 273–283.

RNA capping by mitochondrial and multi-subunit RNA polymerases

Christina Julius , Amber Riaz-Bradley and Yulia Yuzenkova 

Centre for Bacterial Cell Biology, Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne, NE2 4AX, UK

ABSTRACT

Recently, it was found that bacterial and eukaryotic transcripts are capped with cellular cofactors installed by their respective RNA polymerases (RNAPs) during transcription initiation. We now show that mitochondrial RNAP efficiently caps transcripts with ADP – containing cofactors. However, a functional role of universal RNAP – catalysed capping is not yet clear.

ARTICLE HISTORY

Received 14 February 2018
Revised 12 March 2018
Accepted 14 March 2018

KEYWORDS

transcription; RNA polymerase; RNA capping; non-canonical capping; NAD⁺; FAD; dephospho coenzymeA; UDP-GlcNAc; mitochondrial RNA polymerase

The discovery of non-canonical transcript capping

Capping of RNA is no longer seen as an exclusive feature of eukaryotes, thanks to the recent discovery of bacterial transcripts capped by NAD⁺ and 3'-dephosphocoenzyme A (DP-CoA) [1,2]. NAD⁺ is the only cap investigated *in vivo* in *E. coli*, and is found on a number of small RNAs (sRNAs) and messenger RNAs (mRNAs). In addition, a number of currently uncharacterised moieties were found attached to *E. coli* cellular RNA which could potentially also serve as 5' RNA caps [3]. The extent of NAD⁺ modification (NADylation) in the cell varies greatly for different RNA species. The RNA species that are most heavily NADylated *in vivo* [1] are listed on Figure 1A. Even for these species, only a relatively small proportion of the transcripts are capped with NAD⁺ (13% in the case of most heavily NADylated species, namely RNAI – the antisense RNA involved in the regulation of pUC19 plasmid replication [1]). More recently NAD⁺ capping was shown not to be unique for bacteria, as NADylated RNAs were found *in vivo* in *Saccharomyces cerevisiae* and human cells [4,5].



The search for an enzyme that can potentially NADylate RNA transcripts was relatively straightforward, as bacterial RNA polymerase (RNAP) was shown previously to use NAD⁺ as an initiating nucleotide

(given its ADP moiety and free 3' hydroxyl group) [6]. Studies by Bird et al., and Julius and Yuzenkova, using promoter-specific assays, demonstrated that capping can be performed by RNAP on promoters where transcription starts with A [7,8]. These studies showed that the K_m for NAD⁺ in transcription initiation was much lower than the *in vivo* concentration of NAD⁺ (Figure 1B). Furthermore, Bird et al. observed a strong correlation between the extent of NADylation of a chosen transcript *in vivo* and the efficiency of NADylation by RNAP *in vitro* [7]. Eukaryotic RNAPol II was also shown to be able to incorporate NAD⁺, suggesting that the NADylated transcripts observed *in vivo* are also capped by RNAP [7].

Other ADP-containing cofactors were shown to be efficiently incorporated at the 5' end of RNA by RNAP, such as FAD and 3'-dephosphocoenzyme A (but not NADP and NADPH) [7,8]. The efficiency of incorporation for these compounds and their concentration in the cell are lower than those for NAD⁺, suggesting that the possible abundance of these caps is also lower [9].

Cell wall precursors are potentially another class of prokaryotic capping molecules

Dinucleotides UDP-Glucose and UDP-GlcNAc, the precursors of bacterial cell wall synthesis, are even more

CONTACT Yulia Yuzenkova  y.yuzenkova@ncl.ac.uk  Centre for Bacterial Cell Biology, Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne, NE2 4AX, UK

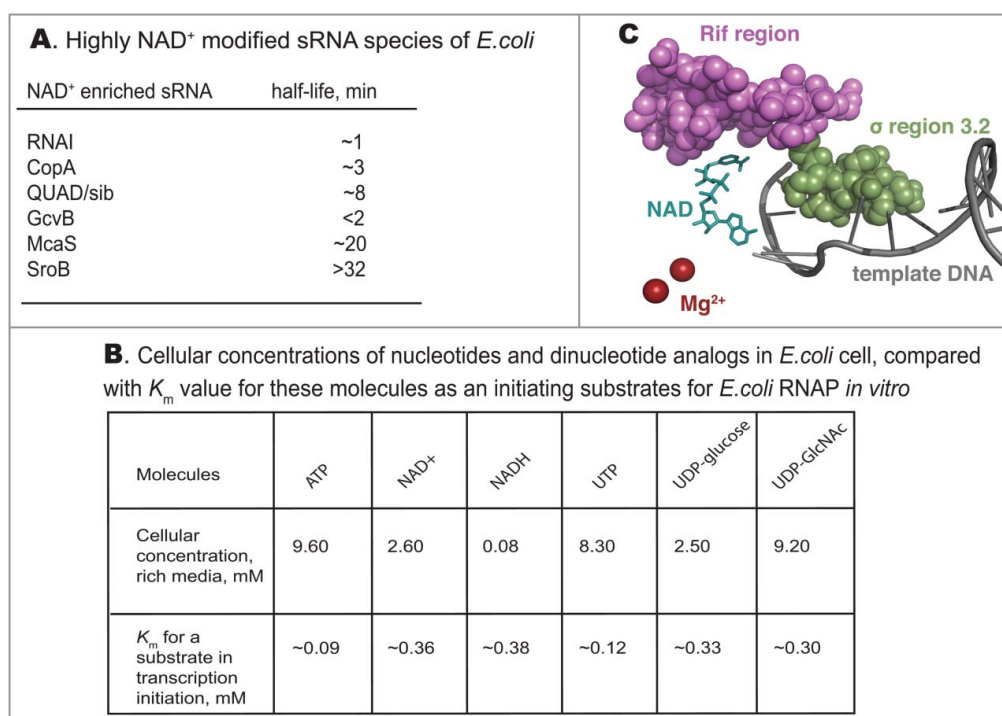


Figure 1. A. A list of 6 heavily NAD⁺ modified RNA species found by Cahova et. al., [1] with half-lives reported in [22]. B. Cellular concentrations of nucleotides and analogs in *E. coli* cell reported by Bennett et. al., [9], and the K_m for their usage as a substrates in transcription initiation [8]. C. Regions of RNAP shown to influence capping efficiency. PDB 5D4D structure of *Thermus thermophilus* RNAP open complex with NADpC was used, NAD is shown in cyan. Part of rifampicin-binding pocket corresponding to cluster I of Rif region of β subunit is in magenta, region 3.2 of σ subunit is in green, template DNA is grey, Mg²⁺ ions are in ruby.

abundant than NAD⁺ in *E. coli* cells grown on rich media (Figure 1B). We recently found that for promoters coding for U at position +1, their RNA transcripts can be efficiently capped *in vitro* by *E. coli* RNA polymerase with UDP-GlcNAc and UDP-Glucose [8]. The relatively low K_m for the incorporation of these substrate at the 5' end of the RNA transcripts by RNAP, favours the probability of *in vivo* capping by UDP-GlcNAc and UDP-Glucose, by analogy with NAD⁺ (Figure 1B). Although less than 10% of *E. coli* promoters code for U at position +1, a link between gene expression and cell wall synthesis could be of potential significance for coordinating biomass and cell wall synthesis.

The ability of RNAP to incorporate variety of known nucleotide-containing molecules at the 5' position of transcript, as well as a number of identified but uncharacterised RNA modifying moieties [3], suggests the existence of a wide repertoire of RNA caps in the cell.

At least two domains of bacterial RNAP determine efficiency of NAD⁺ capping

We showed that initiation with NAD⁺ stabilises short transcripts and favours promoter escape by *E. coli*

RNAP *in vitro* [8]. Whether this stabilisation comes via additional base pairing of cap with the -1 position of the promoter (since NAD⁺ has a nicotine mononucleotide moiety, which may potentially interact with DNA template at -1 position) remains somewhat controversial. Bird et al. showed that the identity of the base at position -1 (-1A vs -1C) affects the efficiency of capping [7]. However, our data suggests that the base at -1 affects initiation in general, without changing the preference for NAD⁺ [8]. Indeed, in the crystal structure of the RNAP initiation complex with a short NADylated transcript, the NMN moiety does not make contacts with DNA but rather faces the protein [7] (Figure 1C). Also, in agreement with the crystal structure, we showed that amino acid changes in the rifampicin-binding pocket of RNAP strongly affected the efficiency of NAD⁺ incorporation, suggesting that observed stabilisation of short capped RNAs is due to interactions between the NAD⁺ cap and the RNAP rifampicin-binding pocket [8] (Figure 1C). Therefore, different configuration of rifampicin-binding pocket may affect NADylation of RNA capping in different bacteria. In contrast to NAD⁺, the incorporation of UDP-containing cell wall precursors was not affected

by the amino acid substitutions in the rifampicin-binding pocket.

Cofactors bound at +1 position may potentially interact with the 3.2 region of initiation factor σ^{70} , which protrudes towards the RNAP active centre [10]. However, we found that a mutant version of σ^{70} lacking region 3.2 ($\sigma^{70\Delta 3.2}$) had no effect on incorporation of NAD⁺, NADH, FAD, UDP-Glucose or UDP-GlcNAc. In contrast, the mature cell wall precursor UDP-MurNAc-pentapeptide, was incorporated by the $\sigma^{70\Delta 3.2}$ mutant of RNAP much more efficiently, suggesting that region 3.2 of σ^{70} may serve to prevent incorporation of advanced cell wall precursors [8] (Figure 1C). Region 3.2 is absent from many sigma factors, suggesting that alternative sigma subunits may allow capping with bulky substrates.

Decapping enzymes for non-canonical caps

The discovery of NudC (NUDIX nicotinamide pyrophosphohydrolase) as an enzyme that removes the NAD⁺ cap from RNA made the parallel between classic eukaryotic and non-canonical capping processes even more striking. *E. coli* NudC was initially described as a housecleaning enzyme hydrolyzing the pyrophosphate bond of NAD⁺/NADH to produce nicotinamide mononucleotide (NMN⁺/NMNH) and AMP [11]. Recently, it was shown that NudC efficiently removes the NAD⁺/NADH cap to produce 5'-monophosphorylated RNA [1]. In eukaryotes, the role of NudC in decapping could be played by NUDIX hydrolases NPY1 in *Saccharomyces cerevisiae* and Nudt19 in *Oryza sativa*, which both showed decapping activity *in vitro* [12]. The activity spectrum of the bacterial NudC is relatively wide; it can remove several ADP analogues from RNA *in vitro*, including DP-CoA [7], consistent with its hydrolase activity towards a broad range of dinucleotides [11].

Removal of cap by NudC was proposed to be the first stage in the degradation of capped RNA to produce a monophosphorylated species, which are a preferred substrate for endonuclease RNaseE [13]. Curiously, however, NudC was not associated with McaS (IsrA) sRNA [14], one of the most highly NADylated sRNAs in *E. coli* (Figure 1A), while other known components of the RNA degradation machinery, such as RNaseE, RNA helicase RhlE and PNPase, were present [14]. This may suggest that the involvement NudC in RNA maturation might be more

complex. Differential decapping by NudC, and its association with target RNAs, could be influenced by the secondary structures of RNAs, as NudC is single-strand dependent [13].

NudC is orthologous to the RppH NUDIX hydrolase, which removes pyrophosphate from triphosphorylated RNA (leaving 5' monophosphate) [15]. A number of additional poorly characterised NUDIX hydrolases in *E. coli* [16] suggests that there might be more potential decapping enzymes for different caps. Notably, neither NudC nor RppH are essential for *E. coli* under normal growth conditions, suggesting possible redundancy of the decapping activities.

Human mitochondrial RNAP efficiently caps RNA with NAD⁺ and other ADP-containing cofactors

Recently, mitochondrial transcripts capped with NAD⁺ were detected in human cells [5]. Mitochondria contain a major cellular pool of NAD⁺ (up to 70%), where it is used for redox reactions and for signalling [17]. We explored the possibility that mitochondrial RNAP (mtRNAP) could cap RNA via a mechanism that is similar to multi-subunit RNAPs. We found that human mtRNAP (hmRNAP) efficiently initiates transcription with NAD⁺, NADH, FAD and DP-CoA on the light strand promoter (LSP; one of the only two human mitochondrial promoters) *in vitro* (Figure 2). The efficiency of initiation with NAD⁺ was approximately 25% compared to ATP, while the other cofactors showed of between 10 to 15%. Our results suggest that mtRNAP is likely to be responsible for adding a NAD⁺ cap to mitochondrial transcripts. Capping in human mitochondria might have consequences for both translation and replication in these organelles. The initially transcribed sequences from both mitochondrial promoters are precursors of tRNAs. It is therefore possible that 5' NADylation might affect their maturation process. Additionally, RNA synthesised from the LSP promoter serves as replication primer [18], and its capping might influence initiation of replication, primer removal and subsequent DNA ligation.

Emerging physiological roles of non-canonical capping

The first experimentally confirmed role for non-canonical NAD⁺ cap in bacteria is an increased

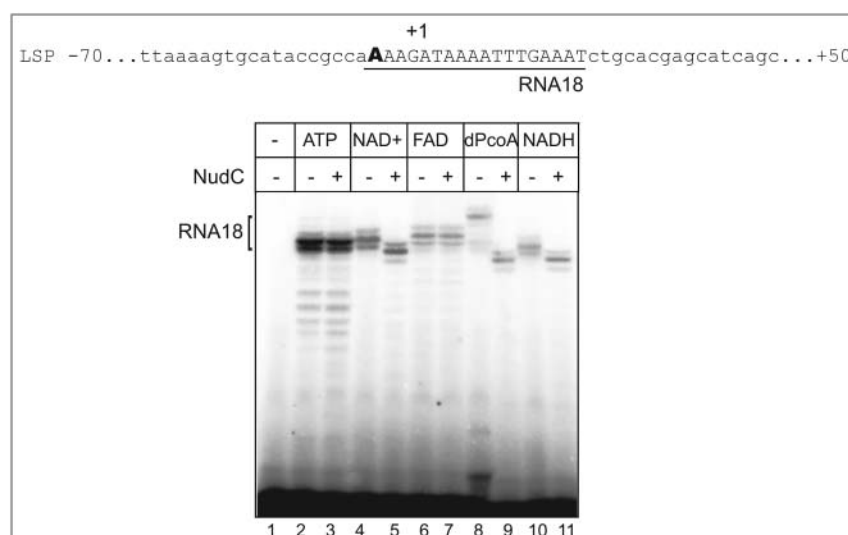


Figure 2. Human mitochondrial RNAP (hmRNAP) incorporates ADP analogs *in vitro*. A partial sequence of the light strand promoter (LSP) is shown, with the initially transcribed sequence underlined. For the assay, 50 nM TFAM, 50 nM hmRNAP, 50 nM TFB2M (purified as described in [23]) were combined with 50 nM of linear DNA fragment containing LSP promoter (positions -70 to +50) in 10 μ l of transcription buffer (40 mM Tris pH 8.0, 10 mM MgCl₂, 10 mM DTT), then ATP or ADP analogs were added to the final concentration of 1mM. Transcription was initiated by the addition of 10 mM MgCl₂, 50 μ M ATP, 300 μ M GTP, 10 μ M [α -³²P]-UTP, 25 Ci/mmol (Hartmann Analytic). After 30 min incubation at 37°C, 500 nM NudC was added to half of the reactions and incubated for additional 15 minutes at 37°C. Transcripts modified with NAD⁺, NADH and DP-CoA (but not ATP or FAD) were susceptible to NudC (lanes 5,9,11) judging from increased mobility of the products. Reactions were stopped by the addition of formamide-containing loading buffer. Products were separated on denaturing polyacrylamide gels (20% acrylamide, 3% bis-acrylamide, 6M urea, 1xTBE), revealed by PhosphorImaging (GE Healthcare), and analysed using ImageQuant software (GE Healthcare).

resistance to degradation, shown for RNAI in the absence of NudC processing [7]. However, this remains controversial, since in other studies [1] deletion of NudC did not affect the overall stability of the RNAI and GcvB populations, the two RNAs most heavily NADylated *in vivo*. Moreover, overall stability of NADylated sRNAs varies widely in wild type *E. coli*, and there is no direct correlation between NADylation and stability (Figure 1A). Notably, in contrast to *E. coli*, NADylation in eukaryotes promotes mRNA decay [5] via decapping by the DXO enzyme, which might additionally supply its 5'-3' degradation activity.

The existence of subpopulations of capped RNA may play a potential role in bistability, the creation of phenotypic variability among clonal population that bacteria use in processes such as dormancy, persistence and sporulation [19]. Capping with dinucleotide analogues might play role in number of regulatory processes involving unstable regulatory RNAs. One example of such process is the type I toxin-antitoxin systems in bacteria, based on translational repression of toxin mRNA by an antisense RNA. This idea is supported by high *in vivo* NADylation of QUAD (sib

RNA (Figure 1A) – antitoxin sRNA preventing the production of the small protein that depolarises the cellular membrane [20].

The extent of capping could be responsive to the changes in cellular metabolism. For example, in *E. coli*, the proportion of NAD⁺ capped RNAI found in stationary phase compared to exponential phase was two-fold higher [1]. Similarly, in yeast, more capped RNA was found in cells grown on synthetic media compared to those grown on the rich media [4]. Since the NAD⁺/NADH balance plays key role in cellular redox homeostasis, capping could connect transcription directly to the cell's redox state. Given the affinity for NAD⁺ is roughly the same as for NADH (Figure 1B), changes in their cellular concentrations will be directly mirrored by the capping of RNA with NAD⁺ or NADH. The functioning of such signalling of course would depend on a mechanism recognising NADylated from NADHylated RNAs. UDP-Glucose and UDP-GlcNAc are the initial substrates for the cascade of reactions leading to the synthesis of cell wall components. It would be tempting to speculate that expression of some cell wall synthesising enzymes could be controlled directly by the pool of

UDP-GlcNAc via capping of +1U transcripts. In general, being rare, RNA modification with cell wall precursors might provide a better regulatory potential, compared to ubiquitous capping with ADP analogs.

Capping might potentially influence translation initiation on a leaderless mRNA.

Another potential cellular role of capping could be the targeting of a specific RNA species, via its cofactor cap, to a protein with affinity for the cognate cofactor, or to a specific subcellular location, *e.g.* to the vicinity of the membrane in the case of UDP-GlcNAc capped RNA.

Intriguingly, we showed that a number of rifampicin resistant RNAPs, including the most widespread clinical isolates, are deficient in capping [8]. This deficiency may contribute to the overall fitness reduction, characteristic for a rifampicin resistant strains [21].

To conclude, despite recent progress, the understanding of non-canonical RNA capping by RNAPs is still patchy. More information is needed to put this type of RNA 5' modification into the category of functional capping, rather than a side reaction of RNAPs. Currently it is hard to envisage a "classic" regulation by the stochastic process of capping. Nevertheless, this "unavoidable" side reaction has to be either used to some advantage or, alternatively, fought against. Both scenarios would have wide ranging cellular consequences with multiple regulatory mechanisms involved. It seems that RNAP has a limited ability to control capping process, apart from alternative sigma factors exchange. It is more feasible to regulate amounts of capped RNA post-transcriptionally, by the linked processes of decapping, alternative folding and RNA chaperons binding. Being a stochastic process, capping might generate variability in a clonal population, which can be exploited at a population level to benefit the organism in adaptation to rapid change in growth conditions. At present, an exact roles of various non-canonical caps in bacteria, eukaryotes and mitochondria are still to be established, as well as full repertoire of enzymes that process non-canonically capped RNAs are to be characterised.

Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.


Acknowledgements


Authors thank Professor Colin Harwood for critical reading of the manuscript.

Funding

This work was supported by a Newcastle University FMS [Studentship] to CJ, a BBSRC [DTP Studentship] to ARB and Royal Society [University Research Fellowship] to YY.

ORCID

Christina Julius  <http://orcid.org/0000-0002-1192-1863>

Yulia Yuzenkova  <http://orcid.org/0000-0003-4036-9235>

References

- [1] Cahova H, Winz ML, Hofer K, et al. NAD captureSeq indicates NAD as a bacterial cap for a subset of regulatory RNAs. *Nature*. 2015;519(7543):374–377. doi:10.1038/nature14020. PMID:25533955
- [2] Kowtoniuk WE, Shen Y, Heemstra JM, et al. A chemical screen for biological small molecule-RNA conjugates reveals CoA-linked RNA. *Proc Natl Acad Sci U S A*. 2009;106(19):7768–7773. doi:10.1073/pnas.0900528106. PMID:19416889; PMCID:PMC2674394.
- [3] Chen YG, Kowtoniuk WE, Agarwal I, et al. LC/MS analysis of cellular RNA reveals NAD-linked RNA. *Nat Chem Biol*. 2009;5(12):879–881. doi:10.1038/nchembio.235. PMID:19820715; PMCID:PMC2842606.
- [4] Walters RW, Matheny T, Mizoue LS, et al. Identification of NAD⁺ capped mRNAs in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A*. 2017;114(3):480–485. doi:10.1073/pnas.1619369114. PMID:28031484; PMCID:PMC5255579
- [5] Jiao X, Doamekpor SK, Bird JG, et al. 5' End Nicotinamide Adenine Dinucleotide cap in human cells promotes RNA decay through DXO-mediated deNADding. *Cell*. 2017;168(6):1015–1027 e10. doi:10.1016/j.cell.2017.02.019. PMID:28283058; PMCID:PMC5371429
- [6] Malygin AG, Shemyakin MF. Adenosine, NAD and FAD can initiate template-dependent RNA synthesis catalyzed by *Escherichia coli* RNA polymerase. *FEBS Lett*. 1979;102(1):51–54. doi:10.1016/0014-5793(79)80926-6. PMID:222618
- [7] Bird JG, Zhang Y, Tian Y, et al. The mechanism of RNA 5' capping with NAD⁺, NADH and desphospho-CoA. *Nature*. 2016;535(7612):444–447. doi:10.1038/nature18622. PMID:27383794; PMCID:PMC4961592
- [8] Julius C, Yuzenkova Y. Bacterial RNA polymerase caps RNA with various cofactors and cell wall precursors. *Nucleic Acids Res*. 2017;45(14):8282–8290. doi:10.1093/nar/gkx452. PMID:28531287; PMCID:PMC5737558
- [9] Bennett BD, Kimball EH, Gao M, et al. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat Chem Biol*. 2009;5(8):593–599. doi:10.1038/nchembio.186. PMID:19561621; PMCID:PMC2754216
- [10] Kulbachinskiy A, Mustaev A. Region 3.2 of the sigma subunit contributes to the binding of the 3'-initiating nucleotide in the RNA polymerase active center and

- facilitates promoter clearance during initiation. *J Biol Chem.* **2006**;281(27):18273–18276. doi:10.1074/jbc.C600060200. PMID:16690607
- [11] Frick DN, Bessman MJ. Cloning, purification, and properties of a novel NADH pyrophosphatase. Evidence for a nucleotide pyrophosphatase catalytic domain in MutT-like enzymes. *J Biol Chem.* **1995**;270(4):1529–1534. doi:10.1074/jbc.270.4.1529. PMID:7829480
- [12] Zhang D, Liu Y, Wang Q, et al. Structural basis of prokaryotic NAD-RNA decapping by NudC. *Cell Res.* **2016**;26(9):1062–1066. doi:10.1038/cr.2016.98. PMID:27561816; PMCID:PMC5034116
- [13] Hofer K, Li S, Abele F, et al. Structure and function of the bacterial decapping enzyme NudC. *Nat Chem Biol.* **2016**;12(9):730–734. doi:10.1038/nchembio.2132. PMID:27428510; PMCID:PMC5003112
- [14] van Nues RW, Castro-Roa D, Yuzenkova Y, et al. Ribonucleoprotein particles of bacterial small non-coding RNA IsrA (IS61 or McaS) and its interaction with RNA polymerase core may link transcription to mRNA fate. *Nucleic Acids Res.* **2016**;44(6):2577–2592. doi:10.1093/nar/gkv1302. PMID:26609136; PMCID:PMC4824073
- [15] Deana A, Celesnik H, Belasco JG. The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal. *Nature.* **2008**;451(7176):355–358. doi:10.1038/nature06475. PMID:18202662
- [16] McLennan AG. The Nudix hydrolase superfamily. *Cell Mol Life Sci.* **2006**;63(2):123–143. doi:10.1007/s00018-005-5386-7. PMID:16378245
- [17] Stein LR, Imai S. The dynamic regulation of NAD metabolism in mitochondria. *Trends Endocrinol Metabol.* **2012**;23(9):420–428. doi:10.1016/j.tem.2012.06.005. PMID:22819213; PMCID:PMC3683958
- [18] Wanrooij PH, Uhler JP, Simonsson T, et al. G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation. *Proc Natl Acad Sci U S A.* **2010**;107(37):16072–16077. doi:10.1073/pnas.1006026107. PMID:20798345; PMCID:PMC2941323
- [19] Veening JW, Stewart EJ, Berngruber TW, et al. Bet-hedging and epigenetic inheritance in bacterial cell development. *Proc Natl Acad Sci U S A.* **2008**;105(11):4393–4398. doi:10.1073/pnas.0700463105. PMID:18326026; PMCID:PMC2393751
- [20] Fozo EM. New type I toxin-antitoxin families from “wild” and laboratory strains of *E. coli*: Ibs-Sib, ShoB-OhsC and Zor-Orz. *RNA Biol.* **2012**;9(12):1504–1512. doi:10.4161/rna.22568. PMID:23182878
- [21] Melnyk AH, Wong A, Kassen R. The fitness costs of antibiotic resistance mutations. *Evol Appl.* **2015**;8(3):273–283. doi:10.1111/eva.12196. PMID:25861385; PMCID:PMC4380921
- [22] Chen H, Shiroguchi K, Ge H, et al. Genome-wide study of mRNA degradation and transcript elongation in *Escherichia coli*. *Mol Syst Biol.* **2015**;11(5):808. doi:10.15252/msb.20159000. PMID:25964259; PMCID:PMC4461401
- [23] Gaspari M, Falkenberg M, Larsson NG, et al. The mitochondrial RNA polymerase contributes critically to promoter specificity in mammalian cells. *EMBO J.* **2004**;23(23):4606–4614. doi:10.1038/sj.emboj.7600465. PMID:15526033; PMCID:PMC533051

FOCUS ARTICLE

Noncanonical RNA-capping: Discovery, mechanism, and physiological role debate

Christina Julius | Yulia Yuzenkova

Centre for Bacterial Cell Biology, Newcastle University, Newcastle upon Tyne, UK

Correspondence

Yulia Yuzenkova, Centre for Bacterial Cell Biology, Newcastle University, Newcastle upon Tyne NE2 4AX, UK.
Email: y.yuzenkova@ncl.ac.uk

Funding information

Royal Society, Grant/Award number: University Research Fellowship; Engineering and Physical Sciences Research Council, Grant/Award number: EP/N031962/1

Recently a new type of 5'-RNA cap was discovered. In contrast to the specialized eukaryotic m⁷G cap, the novel caps are abundant cellular cofactors like NAD⁺. RNAs capped with cofactors are found in prokaryotes and eukaryotes. Unlike m⁷G cap, installed by specialized enzymes, cofactors are attached by main enzyme of transcription, RNA polymerase (RNAP). Cofactors act as noncanonical initiating substrates, provided cofactor's nucleoside base-pairs with template DNA at the transcription start site. Adenosine—containing NAD(H), flavin adenine dinucleotide (FAD), and CoA modify transcripts on promoters starting with +1A. Similarly, uridine-containing cell wall precursors, for example, uridine diphosphate-*N*-acetylglucosamine were shown to cap RNA in vitro on +1U promoters. Noncanonical capping is a universal feature of evolutionary unrelated RNAPs—multisubunit bacterial and eukaryotic RNAPs, and single-subunit mitochondrial RNAP. Cellular concentrations of cofactors, for example, NAD(H) are significantly higher than their K_m in transcription. Yet, only a small proportion of a given cellular RNA is noncanonically capped (if at all). This proportion is a net balance between capping, seemingly stochastic, and decapping, possibly determined by RNA folding, protein binding and transcription rate. NUDIX hydrolases in bacteria and eukaryotes, and DXO family proteins eukaryotes act as decapping enzymes for noncanonical caps. The physiological role of noncanonical RNA capping is only starting to emerge. It was demonstrated to affect RNA stability in vivo in bacteria and eukaryotes and to stimulate RNAP promoter escape in vitro in *Escherichia coli*. NAD⁺/NADH capping ratio may connect transcription to cellular redox state. Potentially, noncanonical capping affects mRNA translation, RNA-protein binding and RNA localization.

This article is categorized under:

RNA Processing > Capping and 5' End Modifications

RNA Export and Localization > RNA Localization

RNA Structure and Dynamics > RNA Structure, Dynamics, and Chemistry

KEYWORDS

dephospho-coenzyme A, FAD, mitochondrial RNA polymerase, NAD⁺, NADH, noncanonical capping, rifampicin, sigma factor, RNA capping, RNA polymerase, transcription, transcription initiation, UDP-GlcNAc

Abbreviations: DP-CoA, dephospho-coenzyme A; FAD, flavin adenine dinucleotide; NAD⁺, nicotinamide adenine dinucleotide; NCIN, noncanonical initiating nucleotide; NMN, nicotinamide mononucleotide; RPo, RNA-promoter open complex; RPc, RNAP-promoter closed complex; TSS, transcription start site; UDP-GlcNAc, uridine diphosphate-*N*-acetylglucosamine; UDP-Glc, uridine diphosphate-glucose; UDP-MurNAc-AA5, uridine diphosphate *N*-acetylmuramic acid pentapeptide

1 | INTRODUCTION—DISCOVERY OF NONCANONICAL CAPPING

In the last decades knowledge of RNA functional roles was extended quite a bit from an information intermediate (mRNA), ribosomal component (rRNA) and peptide formation (tRNA) to regulation (e.g., sRNA and miRNA), immunity and defense (RNA interference), and signaling (e.g., toxin–antitoxin) (Chen, Kowtoniuk, Agarwal, Shen, & Liu, 2009; Liu et al., 2016; Wagner, 2013). Overall, RNA seems to perform a much broader variety of functions than the molecule's sequence and structure conveys; some of these functions rely on RNA modifications. In search for novel RNA modifications, in 2009 the group of David Liu found around 30 unknown small molecule RNA conjugates in *Escherichia coli* (Gram negative) and *Streptomyces venezuelae* (Gram positive) bacterial species (Chen et al., 2009; Kowtoniuk, Shen, Heemstra, Agarwal, & Liu, 2009). Two molecules that were found in both bacteria were dephospho-coenzyme A (DP-CoA and its derivatives) and nicotinamide adenine dinucleotide (NAD⁺). The latter turned out to be present at high abundance as RNA conjugate, comparable in number to an individual tRNA (Chen et al., 2009; Kowtoniuk et al., 2009). Later NAD⁺ and DP-CoA were shown to be attached to the 5'-end of RNA, and a parallel between these modifications and classic eukaryotic m⁷G cap was drawn (Cahova, Winz, Hofer, Nubel, & Jaschke, 2015).

In bacteria, NADylated (NAD-modified) RNA were mainly small regulatory (sRNAs) and 5'-terminal fragments of mRNAs coding for proteins involved in metabolic pathways, stress response control, and for poorly characterized proteins. Notably, different RNAs displayed various degrees of NADylation, always much below 100%. The most frequently NADylated RNA in *E. coli* was found to be RNAI, sRNA for replication control of ColE1 plasmids, with 13% of transcripts bearing NAD⁺. More recently NAD⁺ capped RNAs were found in *Bacillus subtilis*, these were mainly mRNAs (including full length transcripts; Frindert et al., 2018). Corresponding gene products were associated with various cellular pathways, including DNA replication, membrane protein components, sporulation, and oxidation/reduction processes (Frindert et al., 2018).

Following prokaryotes, noncanonical capping of RNA with NAD(H) was also found in eukaryotes. In *Saccharomyces cerevisiae* NADylated RNAs were nuclear mRNAs of genes involved in translation, and mitochondrial RNAs (Walters et al., 2017). In human cells (HEK293T, human kidney tissue) mainly small nuclear (snRNA), and small nucleolar (snoRNA) RNAs were found to be NADylated (Jiao et al., 2017).

As it happens, the newly discovered noncanonical capping of bacterial and eukaryotic RNA challenged the old paradigm and prompted many new questions. First of all, capping of RNA can be viewed no longer as a hallmark of eukaryotic organisms. For a long time it was considered that RNA capping originated from an evolutionary time point after divergence of prokaryotes and eukaryotes. The finding that same noncanonical capping of RNA with cellular cofactors exists in both kingdoms spiked the interest in the discovery in hope to find unknown, ancestral mechanisms and pathways. Conservation of this capping process might imply a functional significance beyond evolutionary ancestry, possibly as previously unrecognized universal mechanism of epigenetic gene expression regulation (Kiledjian, 2018).

The classic cap, methylated guanosine (m⁷G) caters to a sum of functions that are specific to eukaryotes, such as nuclear exit, RNA splicing and assembly of the eukaryotic translation initiation complex (Topisirovic, Svitkin, Sonenberg, & Shatkin, 2011). Potential (common) functions of noncanonical caps in bacteria and eukaryotes are currently actively explored.

2 | THE MECHANISM OF NONCANONICAL CAPPING—INCORPORATION BY RNA POLYMERASE IN TRANSCRIPTION INITIATION

First reports describing the noncanonical capping mechanism were conflicting, as Liu group proposed a posttranscriptional mechanism, similar to classical, eukaryotic capping (Chen et al., 2009; Kowtoniuk et al., 2009). However, other groups proposed transcription initiation (Bird et al., 2016; Julius & Yuzenkova, 2017) as main mechanism (scheme in Figure 1a).

The latter proposition was consistent with a publication about 40 years ago by Malygin and Shemyakin (1979), who used *E. coli* RNA polymerase (RNAP) to promoter-independently synthesize short transcripts initiated with NAD(H) or flavin adenine dinucleotide (FAD). Modern works (Bird et al., 2016; Julius & Yuzenkova, 2017) showed that *E. coli* RNAP initiates transcription with NCINs (noncanonical initiating nucleotides, NAD⁺, DP-coA, and FAD) on a specific promoter with A at +1 position, and where cofactor base-pairs with template with its adenine moiety.

Transcriptional mechanism of noncanonical capping is now established for *E. coli* RNAP, yeast RNAP II (both are multi-subunit RNAPs) as well as T7 and human mitochondrial RNAP (evolutionary unrelated single-subunit enzymes) (Bird et al., 2016; Huang, Bugg, & Yarus, 2000; Julius, Riaz-Bradley, & Yuzenkova, 2018). Given the universality of capping, it is not unreasonable to suggest that all three types of eukaryotic RNAPs (including RNA PolII and PolIII) can incorporate NCINs; the corresponding transcripts were never captured, presumably because they are processed and capped 5' fragments are quickly degraded.

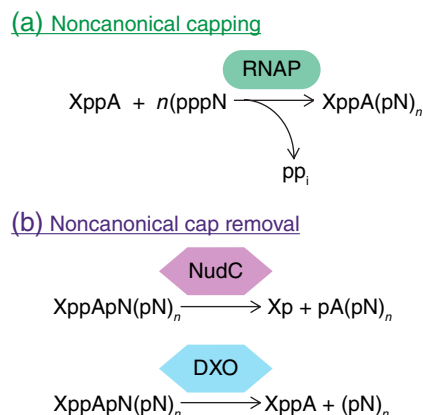


FIGURE 1 Schematic representation of noncanonical capping and decapping. (a) RNA polymerase (RNAP) creates a capped transcript by using the small molecule as initiating substrate instead of canonical nucleoside triphosphate. This small molecule contains a nucleoside diphosphate (here adenosine diphosphate), and another moiety X bound to β -phosphate. RNAP uses incoming nucleoside triphosphate substrates (pppN) to elongate RNA chain, and moiety X is retained as noncanonical cap. (b) Cap removal in bacteria is performed by NUDIX hydrolase NudC, which hydrolyses the phosphate bond between monophosphate-moiety and monophosphate-RNA. DXO decapping enzymes in eukaryotes removes the complete cap (XppA, containing nucleotide moiety of the protein) from the residual RNA molecule. In both cases the product is 5'-monophosphorylated RNA

Figure 2 contains a list of caps with corresponding capping and decapping mechanisms. NAD(H) capping is the most characterized and most widespread, most probably due to a combination of its high availability in different types of cells and high efficiency of incorporation into RNA (see Box 1). K_m to NAD⁺ and NADH in transcription initiation by *E. coli* RNAP (~0.36 and ~0.38 mM, correspondingly) (Julius & Yuzenkova, 2017), are much lower than intracellular concentration, explaining its efficient incorporation into cellular RNA.

Another abundant metabolite in the cell is uridine-diphosphate *N*-acetylglucosamine (UDP-GlcNAc) (Bennett et al., 2009). By analogy to adenine-containing molecules, UDP-GlcNAc as well as another abundant cell wall precursor and metabolite, UDP-glucose (UDP-Glc) were found to initiate transcription on +1U promoters (Julius & Yuzenkova, 2017). K_m of their incorporation by *E. coli* RNAP (~0.33 or ~0.3 mM, respectively) are well below their intracellular concentration in *E. coli* of around 9 mM (Bennett et al., 2009) indicating that capping with cell wall precursors is likely to take place in the cell (Box 1).

3 | A CLOSER LOOK—PROMOTER DNA AND RNAP STRUCTURAL DOMAINS AS DETERMINANTS OF CAPPING EFFICIENCY

Publication of a crystal structure of *Thermus thermophilus* RNAP in complex with template DNA and NADpC short transcript (RNA product of extension of NAD⁺ with CTP) stimulated mechanistic characterization of capping process. Structure of the complex revealed NAD⁺ positioned in vicinity of a template strand of DNA and β -subunit amino acid residues of rifampicin-binding pocket (Rif-pocket; Bird et al., 2016) (Figure 3).

Since NAD⁺ is nucleotide, its NMN moiety can make contacts with -1 position of the template DNA (relative to transcription start site [TSS]). Indeed, base in -1 position of the template was shown to affect efficiency of NAD⁺ incorporation, even though it does not seem to contact NAD⁺ in the crystal (Bird et al., 2016). Presumably during initiation transient base-pairing of template with nicotinamide adenine moiety may occur. The moderate effect of -1 base on the efficiency of NAD⁺ capping was also observed in *B. subtilis* (Frindert et al., 2018).

Recently, Vvedenskaya et al. (2018) described a promoter element around +1A, with consensus sequence H₋₃R₋₂R₋₁A₊₁S₊₂W₊₃W₊₄, which affects capping with NAD⁺ in context of particular *E. coli* promoter (*ptac*, hybrid between *placUV5* and *ptrp* promoter; de Boer, Comstock, & Vasser, 1983). Exchange of each one individual position to the anti-consensus nucleotide (e.g., H to G) results in only 1.2- to 4.1-fold decrease in cofactor incorporation, yet the exchange of whole element to anti-consensus sequence G₋₃Y₋₂Y₋₁A₊₁W₊₂S₊₃S₊₄ decreased likelihood of NAD⁺ incorporation 40-fold. The effect therefore does not seem to be additive, but rather cumulative, suggesting that it is the underlying DNA curvature or duplex stability which affect NAD⁺ incorporation.

Another structural determinant of noncanonical capping in *E. coli* is the Rif-pocket of RNAP. Amino acid changes in Rif-pocket of *E. coli* RNAP show reduced ability to perform NADylation, especially mutation of the residue 516, in close vicinity to NAD⁺ on the crystal structure (Figure 3; Julius & Yuzenkova, 2017). Rifampicin-resistant mutants commonly display a characteristic reduced fitness phenotype (slow growth, increased heat-sensitivity, altered colony morphology) (Brandis, Pietsch, Alemany, & Hughes, 2015). Possibly, the reduced ability to cap RNA might contribute to this phenotype. The Rif-

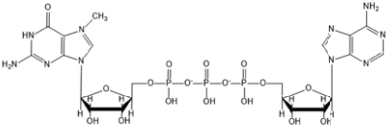
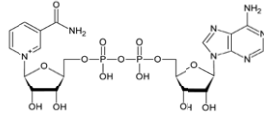
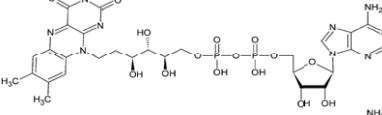
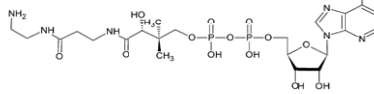
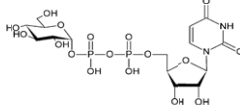
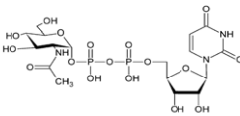
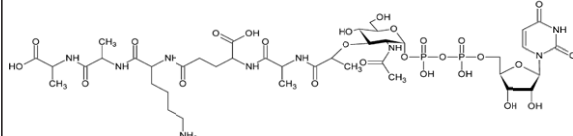
Capping molecules		Capping enzyme	Decapping enzyme
 <p>m⁷G 7-methyl-guanosine, attached to 5' end of RNA</p>	classic cap	phosphatase guanylyltransferase methyltransferase	Dcp2 (NUDIX) DXO
 <p>NAD⁺ nicotinamide adenine dinucleotide</p>  <p>FAD flavine adenine dinucleotide</p>  <p>DP-CoA dephospho coenzyme A</p>	ADP cofactors	bacterial RNAP	NudC (possibly other unknown enzymes)
		eukaryotic RNAPol II	NUDIX enzymes DXO
		mitochondrial RNAP	unknown
 <p>UDP-Glc UDP-glucose</p>  <p>UDP-GlcNAc UDP-N-acetylglucosamine</p>  <p>MurNAc-pentapeptide UDP-N-acetylmuramylpentapeptide</p>	Cell wall synthesis precursors	bacterial RNAP	unknown

FIGURE 2 Comparison of classic and noncanonical capping of RNA

Note. From top to bottom: The classic eukaryotic m⁷G cap is attached co- or posttranscriptionally by a number of specialized enzymes. The process involves dephosphorylation of 5'-triphosphorylated RNA transcript, followed by transferase reaction to attach GTP to the 5'-terminal diphosphate (along with pyrophosphohydrolysis), and finally methylation of guanosine nucleotide. Here, the “inverted” nucleotide serves as the protective structure. Cap removal is performed by NUDIX hydrolases (mainly Dcp2), or DXO proteins. ADP-containing cofactors, here NAD⁺, FAD, and DP-CoA, can be utilized by RNAP in transcription initiation to form the 5'-end of the transcript on promoters where synthesis starts with A (+1 promoters). The rest of the molecule (e.g., NMN of NAD⁺) serves as the protective structure. Likewise, on +1U promoters UDP-containing cell wall precursors, such as UDP-Glc or UDP-GlcNAc, can substitute for UTP in transcription initiation in vitro. The NCIN-caps are removed by NUDIX hydrolases in bacteria, and NUDIX hydrolases or DXO proteins in eukaryotes. RNAP: RNA polymerase

pocket is the site where the nascent transcript passes through the holoenzyme away from the active center of RNAP towards RNA exit channel (Campbell et al., 2001). Presumably interactions of NAD⁺ with amino acid residues of Rif-pocket stabilize capped transcript. This stabilization, along with additional template DNA contacts of NAD⁺, provides plausible explanation for observed positive effect of capping on RNAP escape from RNAI promoter (Julius & Yuzenkova, 2017).

Addition of subinhibitory concentrations of rifampicin to a transcription reaction leads to reduction of RNA elongation (Campbell et al., 2001) and thus promotes release of abortive products. NAD⁺ initiated transcripts are much less sensitive to the same low concentrations of rifampicin on RNAI promoter, suggesting that capped transcript might somewhat compete with the drug for binding in the pocket (Julius & Yuzenkova, 2017). In contrast to *E. coli*, analogous Rif-pocket mutations RNAP in *B. subtilis*, did not affect capping with NAD⁺ in vitro or in vivo, suggesting different conformation of capped transcripts and/or Rif-pocket of *B. subtilis* RNAP (Frindert et al., 2018).

Initiation factor σ^{70} region 3.2 (also called “sigma finger”) protrudes into the RNAP active center during initiation, and is known to affect incorporation of initiating substrates (Kulbachinskiy & Mustaev, 2006). It can perhaps influence incorporation of some NCINs. Indeed, its deletion was shown to enable the incorporation of the bulky uridine-containing cell wall precursor

BOX 1

NAD⁺ concentrations in different cell types and cell compartments

NAD⁺ is likely among the most abundant noncanonical RNA capping molecules, due to its high availability in the cell. In *E. coli* cells grown on glucose NAD⁺ concentration is 2.6 mM (Bennett et al., 2009). At least twice higher NAD⁺ concentrations were reported for *Klebsiella aerogenes*, *Clostridium welchii*, and *Staphylococcus albus* (Wimpenny & Firth, 1972). In yeast values differ between 0.5 and 2.4 mM (Agrimi et al., 2011; Mei & Brenner, 2014). In mammalian cells cytosolic NAD⁺ pool is 0.1–0.8 mM (Canto et al., 2015; Koch-Nolte et al., 2011; Stein & Imai, 2012). Mitochondrial NAD⁺ levels are generally high and in many types of cells they contain major share of NAD⁺ pool, for example, 70% in heart muscle cells and 50% in neural cells (Stein & Imai, 2012). The mammalian nuclear NAD⁺ concentration is lower: 0.07–0.1 mM (Canto et al., 2015; Koch-Nolte et al., 2011).

In comparison, most NAD⁺-utilizing enzymes have a K_m in the micromolar range (Bennett et al., 2009; Canto et al., 2015). For example mammalian lysine deacylases SIRT4 and SIRT5 have 0.035 and 0.98 mM K_m for NAD⁺, correspondingly (Canto et al., 2015). K_m of RNAP for NAD⁺ in transcription initiation is 0.36 mM, in the same range as for metabolic enzymes (Julius & Yuzenkova, 2017).

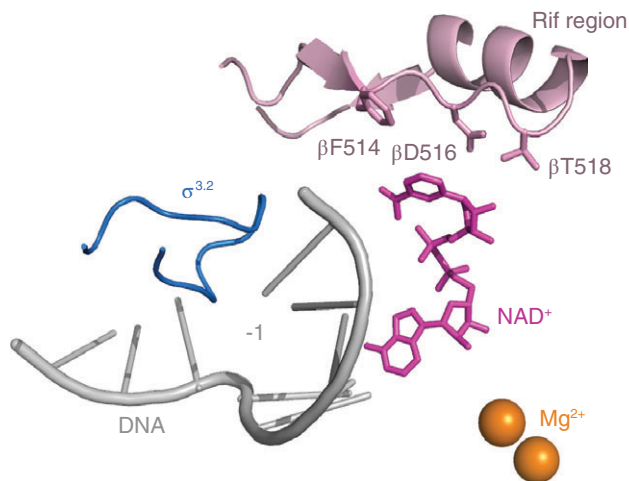


FIGURE 3 Structural features of the capping complex. crystal structure of the initiation complex with NADpC (only nicotinamide adenine dinucleotide [NAD⁺] is shown) product with *Thermus thermophilus* RNA polymerase (RNAP), adapted from PDB ID: 5D4D (Bird et al., 2016). The NAD⁺ (pink) is positioned in vicinity to DNA template (gray) –1 position, RNAP catalytic site (indicated by Mg²⁺ ions), region 3.2 of σ (blue) and part of the rifampicin (Rif)-binding pocket (lilac) of RNAP β subunit. The residues closest to NAD are β F514, β D516, and β T518

UDP-*N*-acetylmuramic acid pentapeptide (UDP-MurNAc-AA5), suggesting that σ finger acts as sterical hindrance for incorporation of large-molecule caps (Julius & Yuzenkova, 2017). Not all σ factors have a region 3.2; thus, alternative σ factors might potentially alter capping activity of the holoenzymes and preference for particular NCINs. Yet, no difference was found between incorporation of NAD⁺, NADH, FAD in vitro by *E. coli* holoenzymes with either housekeeping σ^{70} factor or stationary σ^S (Julius & Yuzenkova, 2017).

No additional transcription factor has yet been described to affect noncanonical capping. Possibly, on promoters sensitive to concentration of initiating substrates, and responsive to ppGpp during formation of transcription-competent complex (Ross et al., 2016; Rutherford et al., 2007), capping could be influenced by DksA protein.

Transcription factor binding, promoter curvature, choice of initiation σ factor, mutations in Rif-pocket, and other promoter and RNAP-features might have a cumulative effect and contribute to variances in proportions of NADylated species for different RNAs. Same factors, combined with differences in cellular concentration, are likely to contribute to the higher copy number of NADylated RNA (>3,000 per cell) as opposed to DP-CoA capped RNA (50–200 copies per cell) (Chen et al., 2009; Kowtoniuk et al., 2009). Yet, currently the extent of noncanonical capping for any transcript cannot be predicted reliably, and overall percentage of capped RNA is net balance between capping and decapping.

4 | MECHANISMS FOR NONCANONICAL CAPS REMOVAL

Classic m⁷G cap is predominantly removed by NUDIX family hydrolase Dcp2. Mammalian genomes contain a high number of NUDIX proteins, and many of them exhibit m⁷G decapping activity: Nudt3 and Nudt16 NUDIX hydrolases decap in vivo and Nudt2, Nudt12, Nudt15, Nudt17, and Nudt19 at least in vitro (Kiledjian, 2018). The NUDIX hydrolases are pyrophosphohydrolases that act on substrates of general structure Nucleoside Diphosphate linked to another moiety X. The NUDIX box common to all NUDIX family members comprises the catalytic amino acids sequence GX₅EX₇REUXEEXGU (Abbondanzieri, Greenleaf, Shaevitz, Landick, & Block, 2005; McLennan, 2006) where the central glutamate residues interact with a divalent cation. Many NUDIX hydrolases possess a zinc-binding motif, which can serve for protein dimerization.

The first enzyme found to remove noncanonical caps was *E. coli* NUDIX hydrolase NudC (Hofer et al., 2016; Jaschke, Hofer, Nubel, & Frindert, 2016), acting on NADylated RNAs (Figure 1b). Previously, it was known as housecleaning enzyme for NAD(H), recycling it to NMN and AMP. It forms a homodimer, and binds one NAD⁺ molecule per monomer as determined by Hofer et al. (2016) and Zhang et al. (2016). The current data on NudC present a conundrum. NudC was shown to act preferentially on capped RNA rather than on NAD(H) substrate in vitro. Also, NudC strongly prefers single-stranded RNA. Yet, mutation studies of NudC demonstrated that it interacts with the NAD⁺-moiety of RNA and not an RNA stretch (Hofer et al., 2016; Zhang et al., 2016) in vitro. This is probably helpful to avoid specific interactions with RNA, but creates a problem in vivo for recognition of capped RNA on a background of NAD(H) excess. Perhaps, by analogy with eukaryotic decapping enzymes, additional adaptor proteins are needed for noncanonical decapping. Another possibility is that NudC acts co-transcriptionally and some yet unknown factors might regulate its activity in connection to RNAP discontinuous movement and RNA synthesis, folding and translation.

The possible number and diversity of NCIN caps is reflected by the finding of a high variety of potential decapping enzymes. *E. coli* encodes 13 different NUDIX hydrolases, many insufficiently characterized. The conserved motif SQPWFPXS classifies most NUDIX hydrolases also as NADH hydrolases (Zhang et al., 2016), thus perhaps a functional overlap exists between different *E. coli* NUDIX hydrolases, and this overlap could account for the moderate increase of NADylation of RNAI found in Δ nudC strain compared to the wild type *E. coli* strain (26% vs. 13%) (Cahova et al., 2015). Among *E. coli* NUDIX enzymes RppH is viewed as decapping enzyme of the sort, since it hydrolyses 5'-triphosphate of RNA producing quickly degradable 5'-monophosphorylated RNA species (Vasilyev & Serganov, 2015).

While there is no NudC ortholog in *B. subtilis*, the RppH ortholog BsRppH was recently shown to have both pyrophosphate removal and NMN-removal activity in vitro (Frindert et al., 2018). Yet, deletion of BsRppH did not affect the level of NADylated RNAs in vivo, suggesting that NADylated RNA is not an in vivo target of BsRppH. Therefore, it is unclear whether the reported effect of BsRppH deletion on expression of about 600 different genes in *B. subtilis* was due to absence of its pyrophosphate—or NMN—removal activity (Frindert et al., 2018).

Besides NUDIX hydrolases, in eukaryotes also the DXO family is known to function as decapping enzymes with activity both on m⁷G cap as well as NAD⁺ cap (Figure 1b). However, DXO cleaves different phosphodiester bond, and removes the complete cap structure as opposed to removal of nicotinamide mononucleotide (NMN) moiety of NAD⁺ by NudC (Jiao et al., 2017) (see Figure 1). Dcp2, Nudt3, and Nudt16 cleave the triphosphate linkage between cap and RNA (Kiledjian, 2018). Thus the mechanisms so far known for mammalian and bacterial NAD-decapping differ. A DXO-equivalent in prokaryotes has not been found to date.

Mitochondrial decapping enzymes have not yet been reported, although at least six NUDIX hydrolases are targeted to this organelle in mammals (Mildvan et al., 2005). One of them, Nudt13 has a strong affinity to NADH and NADPH, and might act as de-capping enzyme for NADylated mitochondrial transcripts (Abdelraheim, Spiller, & McLennan, 2017).

5 | POTENTIAL FUNCTIONAL IMPLICATIONS OF NONCANONICAL CAPPING

5.1 | NCIN capping and RNA stability—up or down regulation?

The first experimentally confirmed global in vivo role for noncanonical NAD⁺ cap so far is its effect on RNA stability. This effect is strikingly opposite for bacteria and eukaryotes—in *E. coli* NAD⁺ cap increases RNAI stability (in the absence of NudC; Bird et al., 2016), in *S. cerevisiae* NADylation promotes mRNA decay via decapping and 5'–3' degradation activity of DXO enzyme (Jiao et al., 2017).

We expect, perhaps intuitively, these effects to be similar, and the cap to play a universal protective role. Probably difference between kingdoms might actually stem from different experimental conditions, that is, crucial absence of decapping enzyme, NudC in case of bacteria, leading to specific stabilization of capped RNA. Notably, in work of Cahova et al. (2015)

TABLE 1 Examples of sRNAs highly NADylated in vivo, sorted by half-life

sRNA	<i>cis</i> or <i>trans</i>	Target mRNA(s)	Protein binding	Half-life (min)	Capping-level in vivo (%)	Function	References
RNAI	<i>cis</i>	RNA2	Rop	1	13	ColE1 replication control	Cahova et al. (2015) and Tamm and Polisky (1985)
GlmY	<i>trans</i>	GlmZ	PAPI	1.4	1.6	Peptidoglycan synthesis pathway	Cahova et al. (2015), Reichenbach et al. (2008), and Vvedenskaya et al. (2018)
GcvB	<i>trans</i>	CycA	CsrA	2	2	Peptide/aa transport, glycine transport	Cahova et al. (2015), Pulvermacher, Stauffer, and Stauffer (2008), and Vvedenskaya et al. (2018)
CopA	<i>trans</i>	CopT	n.d.	3	n.d., highly enriched for NAD ⁺	r1 plasmid replication	Cahova et al. (2015) and Gerhard, Wagner, and Nordstrom (1986)
QUAD/sib	<i>cis</i>	ibs	n.d.	8	n.d., highly enriched for NAD ⁺	Antotoxin repressing Ibs toxic protein production	Cahova et al. (2015) and Fozo (2012)
McaS	<i>trans</i>	YdeH, YdeT	Hfq and CsrA	20	5.1	Biofilm regulation	Cahova et al. (2015), Jorgensen et al. (2013), and Vvedenskaya et al. (2018)
DsrA	<i>trans</i>	RpoS, Hns, MreB, RbsD	Hfq	23	n.d., highly enriched for NAD ⁺	Transcription, cell wall synthesis, ribose metabolism	Cahova et al. (2015), Cayrol et al. (2009) and Wu et al. (2017)
ChiX	<i>trans</i>	ChiP	Hfq	27	1.6	Regulation of outer membrane channel	Cahova et al. (2015), Edwards et al. (2011), Mandin and Gottesman (2009), and Vvedenskaya et al. (2018)
GadY	<i>trans/cis</i> ^a	GadX, GDS	Hfq/Hfq-independent ^a	n.d.	3.9	pH stress response	Bird et al. (2016), Cahova et al. (2015), Negrete and Shiloach (2015), Opdyke et al. (2011), and Vvedenskaya et al. (2018)

Note. Stability of RNA does not seem to correlate with NADylation percentage (cells shading reflects the magnitude of these parameters). NADylated transcripts were detected either by NAD-captureSeq (Cahova et al., 2015) or CapZyme-Seq (Vvedenskaya et al., 2018). Percentage of capping was calculated by Vvedenskaya et al. (2018) (except for RNAI). Highly enriched RNA species were found by Cahova et al. (2015) (exact percentage of NADylation is not known). A further feature is interaction with RNA chaperones, such as Hfq, or other protein partners. n.d.: not determined.

^a Conflicting data.

NudC deletion did not affect the overall stability of RNAI and GcvB populations, two RNAs most heavily NADylated in vivo. In general, overall stability of NADylated sRNAs varies widely in wild type *E. coli* (see Table 1).

The nature and structure of the 5'-end of RNA plays crucial role in RNA turnover in both bacteria and eukaryotes. Most eukaryotic RNases are 5'-3'-exonucleases. Therefore, one of the first steps of eukaryotic mRNA degradation is decapping (Kushner, 2004).

Although the main degradation enzyme in *E. coli* RNase E is an endonuclease, it has a preference for RNA substrates with 5'-monophosphate ends. Therefore, capped RNA degradation might rely on processing with NUDIX hydrolases to yield a monophosphorylated 5'-end. Similarly, *B. subtilis* endo- and 5'-3'-exonuclease RNase J1 demonstrated a preference for 5'-monophosphorylated RNA (Yao, Sharp, & Bechhofer, 2009). NAD⁺ cap inhibits RNase J1-mediated degradation to a similar extent as the 5'-triphosphate (Frindert et al., 2018).

In *E. coli*, the main breakdown machinery for RNA forms a degradosome (Mohanty & Kushner, 2016; Silva et al., 2011). This multiprotein complex comprises main endoribonuclease RNase E, exoribonuclease and polyadenylation enzyme PNPase, RNA helicases, and enolase. However, neither of prokaryotic decapping enzymes—RppH nor NudC are associated with degradosome. Perhaps, decapping enzyme(s) interact only transiently with degradosome. The finding that most bacterial NADylated transcripts besides sRNAs are 5'-fragments of mRNAs (Cahova et al., 2015) might imply that the found mRNA

fragments are degradation intermediates, thus degradation would have started despite the presence of cap. This could indicate that at least in some cases noncanonical cap in bacteria can promote degradation, just like in eukaryotes (Kiledjian, 2018). Possibly, the cap leads to a yet undefined degradation pathway in bacteria. Alternatively, found fragments might belong to RNA substrates which RNaseE processes via alternative, 5' independent "direct entry" mechanism (Bouvier & Carpoussis, 2011).

Capping might affect RNA stability indirectly via polyadenylation. While polyadenylation in eukaryotes serves as a stabilizing structure, in prokaryotes the addition of the poly(A) tail initiates RNA degradation (Kushner, 2004). Likely, all bacterial RNAs could be polyadenylated at some point (Sarkar, 1996). Due to main poly(A)-polymerase PAPI preference for monophosphorylated substrates (Feng & Cohen, 2000), NCIN-capping might inhibit polyadenylation. Moreover, polyadenylation inhibition might further decrease efficiency of degradation since poly(A) tail is required for RNA 3'–5'-degradation (Mohanty & Kushner, 2016). Furthermore, the polyadenylation state of RNA affects its regulation by sRNAs. For example, RNA chaperone Hfq pairs *trans*-acting sRNAs with their target mRNAs, and has a preference for binding of poly(A) tail structures (Mohanty & Kushner, 2016; Silva et al., 2011; Wagner, 2013). Therefore, repression of polyadenylation by NCIN-caps could modulate effectivity of sRNA regulation. Overall, the net result on RNA stability of noncanonical capping and polyadenylation combination might vary for different RNA species.

It is possible that noncanonical decapping processes are influenced by modifications on adjoining base(s) of RNA, just like the m⁷G cap. 30% of eukaryotic cellular RNA carries methyl group on the first nucleotide of the transcript adjacent to the cap, and many transcripts are di- and tri-methylated. On methylated RNAs noncanonical cap might resist de-capping by NUDIX hydrolase Dcp2, similarly to the classical cap (Kiledjian, 2018; Mauer et al., 2017).

5.2 | Potential effect of noncanonical capping on translation initiation

Significant number of NADylated mRNAs was identified in *E. coli*, *B. subtilis*, and *S. cerevisiae* (Cahova et al., 2015; Frindert et al., 2018; Walters et al., 2017). The eukaryotic m⁷G cap is an important structural feature of assembly of the translation initiation complex (Ramanathan, Robb, & Chan, 2016). Notably m⁷G is replaceable by artificially generated alternative cap structures (e.g., m⁷Gp₃m⁷G) some of which can inhibit or promote translation initiation (Grudzien et al., 2004). It was shown that different GTP analogs attached during classic eukaryotic capping *in vitro* were more stable than uncapped RNA after transfection into HEK293 cells. However, most of the unnatural caps were not recognized by the translation machinery (Issur, Bougie, Despains, & Bisailon, 2013). Apparently, in eukaryotic translation NAD⁺ cap does not function as classic m⁷G cap. Experiments with luciferase reporter gene in cell culture have revealed that presence of NAD⁺-cap results in translation levels similar to uncapped (5'-triphosphorylated) RNA (Jiao et al., 2017).

How noncanonical capping affects translation of mRNAs in bacteria is yet unknown. Most probably it does not affect translation initiation on mRNAs with ribosomal binding sites downstream of TSS. On the other hand, regulation via caps at the 5'-end of leaderless mRNA, in other words, with NAD⁺ UG starting codon instead AUG, is very possible. On leaderless RNA, the AUG codon is the necessary and sufficient recognition signal for ribosome binding (in complex with IF2 and fMet-tRNA) which can lead to successful translation initiation (Moll, Grill, Gualerzi, & Blasi, 2002). It remains to be determined whether capping stimulates or inhibits ribosome binding, or translation initiation on capped RNA relies on the presence of additional factors. Translation regulation by noncanonical capping in bacteria is likely, since many bacteria encode high numbers of leaderless transcripts, for example, one third of mycobacterial transcripts are leaderless (Shell et al., 2015). The same is relevant for mitochondria, where mRNAs are generally leaderless (Jones, Wilkinson, Hung, Weeks, & Spremulli, 2008).

5.3 | Capping and regulatory sRNAs functions of bacteria

sRNAs are the most frequently NADylated RNAs in *E. coli* (Cahova et al., 2015; Vvedenskaya et al., 2018). Table 1 sums up known half-lives, interaction partners and functions of several sRNAs which were shown to be highly NADylated in *E. coli*. Percentage of NADylated species, if known, is from Cahova et al. (2015) and Vvedenskaya et al. (2018). These RNAs do not seem to share many common features in terms of their half-life, *cis* or *trans* mechanism of action (see below) or their ability to bind RNA chaperones, such as Hfq.

Cis-acting sRNAs are encoded on the anti-sense strand of their target mRNA and have complete sequence homology, thus they are specific to their particular mRNA. This type of regulation is often found to control copy number of plasmids or transposons (Waters & Storz, 2009). One example for the *cis*-acting sRNA is RNAI, the most highly NADylated RNA in *E. coli* (Cahova et al., 2015). RNAI regulates the replication of plasmid ColE1 by sequestering the ColE1 replication primer (Tamm & Polisky, 1985).

Many bacteria use regulatory toxin-antitoxin systems. They consist of a stable toxin protein, and an unstable antitoxin. In Type I systems antitoxin is sRNA acting in antisense to repress translation of toxin mRNA (Berghoff & Wagner, 2017). Both antitoxin sRNA and toxin mRNAs stability can be affected by noncanonical capping. Indeed, one of the highly NADylated RNA examples in Table 1 is QUAD (sib) sRNA—antitoxin sRNA preventing expression of small protein depolarizing cellular membrane (Fozo, 2012).

Trans-acting sRNAs are generally transcribed from another location than their target mRNA. The sequence homology is limited and many *trans* sRNAs have several target RNAs. RNA chaperone Hfq frequently mediates between sRNA and target mRNA. *Trans* sRNAs act as translational silencers, binding and occluding the RBS or other regions in the 5'-UTR of the mRNA (Waters & Storz, 2009). One example of a *trans*-acting sRNA is GcvB, shown to be highly NADylated in vivo (Cahova et al., 2015). It binds mRNA upstream of the translation start site, resulting in degradation of dsRNA (Waters & Storz, 2009). *Trans* sRNAs are often regulated by a specific transcription factor in response to the environmental conditions (e.g., GcvB transcription factor GcvA is part of elevated glycine response), and they pass on this response by translation control of their various targets (Waters & Storz, 2009).

It was shown (Malecka, Strobecka, Sobanska, & Olejniczak, 2015) that interaction with RNA chaperone Hfq is influenced by the 5'-terminus of sRNA. Hfq can form a complex with PNPase, PAPI, and RNase E for the purpose of maturation or degradation of polycistronic transcripts (Mohanty & Kushner, 2016). Interestingly, many examples of sRNAs that bind Hfq strongly and thus have the tendency to displace more weakly bound sRNAs on Hfq, are also among the most frequently NADylated sRNAs identified (Cahova et al., 2015), for example ChiX and DsrA. Notably, ChiX is a very stable RNA (Table 1), its features that confer resistance to degradation are yet unknown; to date ChiX is the only known prokaryotic catalytic sRNA (Mandin & Gottesman, 2009). Another frequently NADylated sRNA, McaS, binds to Hfq, as well as to global RNA-binding protein CsrA (Malecka et al., 2015). Table 1 describes some further examples of NADylated sRNAs and their function and interaction partners.

We can hypothesize that other regulatory functions of RNA could be modulated by NCIN-capping. For example, “RNA thermometers” and riboswitches (*cis*-acting regulatory elements in the 5'-UTR that bind ligands) assume a secondary structure in response to environmental stimulus, can bind ligands and modulate translation (Waters & Storz, 2009). Riboswitches form large complex secondary structures, the 5'-terminus can be either included in this structure, or remains single stranded, contributing to the “on” or “off” conformation of the switch (Montange & Batey, 2008). It would be interesting to see if/how non-canonical capping might affect regulation. Perhaps, the formation of a dsRNA region could be affected by presence of the cofactor moiety, possibly destabilizing/stabilizing one of the alternative conformations. Also, presence of a nonnucleotide moiety in vicinity to the riboswitch ligand binding site could affect its binding (Montange & Batey, 2008).

5.4 | Noncanonical cap-dependent RNA localization

Another potential outcome of noncanonical capping could be RNA targeting via its cap to a specific protein with affinity for the cognate cofactor or to a specific subcellular location. For a long time it was thought that mRNA co-localizes with the bacterial nucleoid, since translation is coupled with transcription (Vogel & Jensen, 1994). Yet more recently it was found that indeed only 4% of translation complexes are coupled and co-localized with transcription (Buskila, Kannaiah, & Amster-Choder, 2014). Moreover, a number of mRNAs in *E. coli*, independently from their translation, were found localized to intracellular sites of their encoded proteins (Nevo-Dinur, Nussbaum-Shochat, Ben-Yehuda, & Amster-Choder, 2011).

The mechanism of mRNA localization is unknown, and simple diffusion through the cytoplasm is unlikely the only mode of transport for RNAs, which can be bulky (especially in complex with RNA-binding proteins). In eukaryotes, RNA is believed to display *cis*-acting localizing elements, or “zip-codes” in untranslated regions. Their mechanism of action is unresolved (Buskilay et al., 2014). In light of mRNA colocalization with proteins, the case of RNA capped with cell wall precursors (Julius & Yuzenkova, 2017) is potentially interesting. UDP-Glc and UDP-GlcNAc are present at high concentrations in the cell and are turned over frequently (Konopka, 2012). It is possible that diffusion of these molecules throughout the cell facilitates their incorporation into RNA, and their association to the cell wall and interaction with proteins in the cell periphery might transport or “tag” the transcript towards the bacterial inner membrane.

5.5 | Protein-NCIN capped RNA interactions

Besides RNA-binding proteins like RNases, Hfq, and CsrA, many protein enzymes involved in metabolism interact with RNA. Also, nucleoside containing molecules such as the adenine-containing cofactors (NAD⁺, NADH, FAD, DP-CoA, and more) are often bound by enzymes of metabolism. In eukaryotes, growing number of metabolic enzymes are found to “moonlight” as RNA-binding proteins (Castello, Hentze, & Preiss, 2015). It is tempting to speculate that at least some of these protein-RNA interactions happen with assistance of a noncanonical cofactor cap.

5.6 | Noncanonical capping and metabolism

Since NAD^+/NADH balance reflects cellular redox homeostasis, capping could connect transcription directly to redox state, as affinity in transcription for NAD^+ is roughly the same as for NADH (Julius & Yuzenkova, 2017).

In addition, the overall extent of capping depends on metabolic state of the cell. For example, in stationary phase *E. coli*, two-fold higher proportion of NAD^+ -capped RNA was found (Bird et al., 2016). Likewise, in yeast grown on synthetic media more RNA molecules are capped compared to the rich one (Chen et al., 2009).

Similarly, cell wall metabolism could potentially affect transcription on +1U promoters. UDP-Glc and UDP-GlcNAc are precursors for peptidoglycan synthesis in bacteria. Their incorporation as NCIN-cap by bacterial RNAP was shown in vitro, and might be responsive to the concentration of those molecules in vivo. Furthermore, pool of UDP-GlcNAc might control expression of some cell-wall-making enzymes via capping of +1U transcripts. Inhibition of cell wall synthesis by antibiotics would increase the precursor pool and probably provide feedback for cell wall making enzymes expression.

5.7 | Alternative model for noncanonical capping: Post-transcriptional

When the first noncanonical caps DP-CoA and NAD^+ were identified in 2009 (Chen et al., 2009; Kowtoniuk et al., 2009), the authors proposed a posttranscriptional mechanism for attachment of NAD^+ and DP-CoA caps. Later transcription initiation was identified as a major noncanonical capping mechanism. However, the possibility of an additional mechanism involving enzymes other than RNAPs, should not be entirely dismissed. Jiao et al. (2017) proposed that at least in mammalian cells, an alternative “ NAD^+ -capping” mechanism occurs. Their hypothesis was fuelled by the finding that small nucleolar RNAs (snoRNAs) and small Cajal body RNAs (scaRNAs) frequently carry an NAD^+ modification. These RNAs are susceptible to DXO-mediated “ NAD^+ -capping” and predominantly generated by processing of a transcript. However, at least snoRNAs are known to be transcribed in various ways, and some are transcribed from an independent promoter, rather than generated via processing (Dieci, Preti, & Montanini, 2009). Therefore, the RNAP might still be responsible for capping of these RNAs. Nevertheless, we should keep in mind that the view that RNAP incorporation is the main if not the only capping mechanism is based on the observation that an extent of in vivo capping correlates with in vitro efficiency of NAD^+ incorporation by RNAP in just several cases (Bird et al., 2016). Therefore, currently we do not have any basis to convincingly rule out a transcription-independent capping mechanism.

Intiguingly, protein-independent mechanism for producing RNAs with cofactors attached to 5'-end exists (Huang et al., 2000). Ribozymes able to make NAD^+ , FAD, and CoA capped RNA, by attaching small molecule precursors to a 5'-terminal ATP were evolved in vitro. This ribozyme activity is consistent with the conservation of the nucleotide containing coenzymes throughout all types of cells, since cofactors might be viewed as evolutionary remnants of “RNA world” (White, 1976). Hypothetically, the ribozyme reactions might once have been the main pathway, and could also today still be working unrecognized in in vivo synthesis of capped RNA.

6 | CONCLUSION: NONCANONICAL CAPPING—STOCHASTIC EVENT OR REGULATED PROCESS?

The very universality of noncanonical capping—an ability of unrelated multi- and single-subunit RNAPs to incorporate NCINs efficiently, poses a question whether this type of capping is just a stochastic process, side reaction of transcription stemming from relaxed substrate specificity of RNAPs for initiating substrate. Kiledjian (2018) argues that RNAP infidelity of incorporation could not account for the observation that different RNAs show different levels of NAD^+ -capping. Yet, based on currently available data it is hard to envisage mechanism for capping regulation at the stage of NCINs incorporation, especially given the considerable excess of substrates (see Box 1) in comparison to K_m for them in transcription initiation, and weak consensus for promoters (i.e., lack of specific features) for NAD^+ incorporation, at least in *E. coli* (Veedenskaya et al., 2018). Kiledjian (2018) proposes to vary cellular NAD^+ concentrations to determine if NAD^+ incorporation is a stochastic or a specific event. The verification of functional roles of the caps in vivo is likely the only way to prove that NCIN-capping is a regulatory event and not a simple side reaction of transcription.

Nevertheless, even if NCINs incorporation is entirely stochastic event, cells have to deal with it—use it or fight against it.

Universality of noncanonical capping is mirrored by widespread occurrence of NUDIX and DXO families of (potential) decapping proteins in prokaryotes and eukaryotes. Members of DXO decapping family are involved in general 5'-end RNA quality control (Kiledjian, 2018). Likewise, NUDIX family proteins were described as housecleaning enzymes, displaying broad substrate range (McLennan, 2013). Many of them show activity as antimutator enzymes (removal of mutagenic oxidized radicals of nucleotides) leading to hypothesis that this activity might have been the original function of NUDIX hydrolases before cleavage of nucleotide analogs, cofactors, and other molecules evolved (McLennan, 2013).

The efficiency of decapping could be predictably influenced by RNA secondary structure, binding of chaperons and other factors. Therefore, decapping stage might be a true regulatory point and a process responsible for overall extent of RNA modification.

Research into noncanonical capping in prokaryotes and eukaryotes is still in its infancy; future work will hopefully answer the following questions:

- What is the full repertoire of noncanonical caps?
- How different is the spectrum of noncanonical capping in various prokaryotic and eukaryotic species (and their organelles)?
- Are there any other polymerases initiating with NCINs?
- Are there any proteins specifically recognizing noncanonically capped RNA?
- Does posttranscriptional, RNAP independent mechanism for noncanonical capping exist?
- Finally, perhaps the most crucial question: what are the physiological roles of noncanonical capping?

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

RELATED WIREs ARTICLE

[Translation initiation by cap-dependent ribosome recruitment: Recent insights and open questions](#)

FURTHER READING

Bochkareva, A., Yuzenkova, Y., Tadigotla, V. R., & Zenkin, N. (2012). Factor-independent transcription pausing caused by recognition of the RNA-DNA hybrid sequence. *The EMBO Journal*, 31, 630–639.

REFERENCES

- Abbondanzieri, E. A., Greenleaf, W. J., Shaeviz, J. W., Landick, R., & Block, S. M. (2005). Direct observation of base-pair stepping by RNA polymerase. *Nature*, 438, 460–465.
- Abdelrahman, S. R., Spiller, D. G., & McLennan, A. G. (2017). Mouse Nudt13 is a mitochondrial NUDIX hydrolase with NAD(P)H pyrophosphohydrolase activity. *The Protein Journal*, 36, 425–432.
- Agimi, G., Brambilla, L., Frascotti, G., Pisano, I., Porro, D., Vai, M., & Palmieri, L. (2011). Deletion or overexpression of mitochondrial NAD⁺ carriers in *Saccharomyces cerevisiae* alters cellular NAD and ATP contents and affects mitochondrial metabolism and the rate of glycolysis. *Applied and Environmental Microbiology*, 77, 2239–2246.
- Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., Van Dien, S. J., & Rabinowitz, J. D. (2009). Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nature Chemical Biology*, 5, 593–599.
- Berghoff, B. A., & Wagner, E. G. H. (2017). RNA-based regulation in type I toxin-antitoxin systems and its implication for bacterial persistence. *Current Genetics*, 63, 1011–1016.
- Bird, J. G., Zhang, Y., Tian, Y., Panova, N., Barvik, I., Greene, L., ... Nickels, B. E. (2016). The mechanism of RNA 5' capping with NAD⁺, NADH and desphospho-CoA. *Nature*, 535, 444–447.
- Bouvier, M., & Carpousis, A. J. (2011). A tale of two mRNA degradation pathways mediated by RNase E. *Molecular Microbiology*, 82, 1305–1310.
- Brandis, G., Pietsch, F., Alemayehu, R., & Hughes, D. (2015). Comprehensive phenotypic characterization of rifampicin resistance mutations in salmonella provides insight into the evolution of resistance in *Mycobacterium tuberculosis*. *The Journal of Antimicrobial Chemotherapy*, 70, 680–685.
- Buskila, A. A., Kannaiah, S., & Amster-Choder, O. (2014). RNA localization in bacteria. *RNA Biology*, 11, 1051–1060 <https://doi.org/10.4161/rna.36135>
- Cahova, H., Winz, M. L., Hofer, K., Nubel, G., & Jaschke, A. (2015). NAD captureSeq indicates NAD as a bacterial cap for a subset of regulatory RNAs. *Nature*, 519, 374–377.
- Campbell, E. A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., & Darst, S. A. (2001). Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell*, 104, 901–912.
- Canto, C., Menzies, K. J., & Auwerx, J. (2015). NAD(+) metabolism and the control of energy homeostasis: A balancing act between mitochondria and the nucleus. *Cell Metabolism*, 22, 31–53.
- Castello, A., Hentze, M. W., & Preiss, T. (2015). Metabolic enzymes enjoying new partnerships as RNA-binding proteins. *Trends in Endocrinology and Metabolism*, 26, 746–757. <https://doi.org/10.1016/j.tem.2015.09.012>
- Cayrol, B., Geinguenaud, F., Lacoste, J., Busi, F., Le Dérout, J., Piétrement, O., ... Arluison, V. (2009). Auto-assembly of *E. coli* DsrA small noncoding RNA: Molecular characteristics and functional consequences. *RNA Biology*, 6, 434–445. <https://doi.org/10.4161/rna.6.4.8949>
- Chen, Y. G., Kowtoniuk, W. E., Agarwal, I., Shen, Y., & Liu, D. R. (2009). LC/MS analysis of cellular RNA reveals NAD-linked RNA. *Nature Chemical Biology*, 5, 879–881.
- de Boer, H. A., Comstock, L. J., & Vasser, M. (1983). The tac promoter: A functional hybrid derived from the trp and lac promoters. *Proceedings of the National Academy of Sciences of the United States of America*, 80, 21–25.
- Dieci, G., Preti, M., & Montanini, B. (2009). Eukaryotic snoRNAs: A paradigm for gene expression flexibility. *Genomics*, 94, 83–88.
- Edwards, A. N., Patterson-fortin, L. M., Vakulskas, C. A., Jeffrey, W., Potrykus, K., Vinella, D., ... Babitzke, P. (2011). Circuitry linking the Csr and stringent response global regulatory systems. *Molecular Cell*, 80, 1561–1580. <https://doi.org/10.1111/j.1365-2958.2011.07663.x>

- Feng, Y., & Cohen, S. N. (2000). Unpaired terminal nucleotides and 5' monophosphorylation govern 3' polyadenylation by *Escherichia coli* poly(A) polymerase I. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 6415–6420.
- Fozo, E. M. (2012). New type I toxin-antitoxin families from “wild” and laboratory strains of *E. coli*: Ibs-Sib, ShoB-OhsC and Zor-Orz. *RNA Biology*, 9, 1504–1512.
- Frindert, J., Zhang, Y., Nubel, G., Kahloon, M., Kolmar, L., Hotz-Wagenblatt, A., ... Jaschke, A. (2018). Identification, biosynthesis, and decapping of NAD-capped RNAs in *B. subtilis*. *Cell Reports*, 24, 1890–1901.e8.
- Gerhard, E., Wagner, H., & Nordstroem, K. (1986). Structural analysis of an RNA molecule involved in replication control of plasmid R1. *Nucleic Acids Research*, 14, 8919–8932. <https://doi.org/10.1093/nar/gkn907>
- Grudzien, E., Stepinski, J., Jankowska-Anyszka, M., Stolarski, R., Darzynkiewicz, E., & Rhoads, R. E. (2004). Novel cap analogs for in vitro synthesis of mRNAs with high translational efficiency. *RNA*, 10, 1479–1487. <https://doi.org/10.1261/rna.7380904>
- Hofer, K., Li, S., Abele, F., Frindert, J., Schlotthauer, J., Grawenhoff, J., ... Jaschke, A. (2016). Structure and function of the bacterial decapping enzyme NudC. *Nature Chemical Biology*, 12, 730–734.
- Huang, F., Bugg, C. W., & Yarus, M. (2000). RNA-catalyzed CoA, NAD, and FAD synthesis from phosphopantetheine, NMN, and FMN. *Biochemistry*, 39, 15548–15555.
- Issur, M., Bougie, I., Despins, S., & Bisailon, M. (2013). Enzymatic synthesis of RNAs capped with nucleotide analogues reveals the molecular basis for substrate selectivity of RNA capping enzyme: Impacts on RNA metabolism. *PLoS One*, 8, e75310.
- Jaschke, A., Hofer, K., Nubel, G., & Frindert, J. (2016). Cap-like structures in bacterial RNA and epitranscriptomic modification. *Current Opinion in Microbiology*, 30, 44–49.
- Jiao, X., Doamekpor, S. K., Bird, J. G., Nickels, B. E., Tong, L., Hart, R. P., & Kiledjian, M. (2017). 5' end nicotinamide adenine dinucleotide cap in human cells promotes RNA decay through DXO-mediated deNADding. *Cell*, 168(1015–1027), e1010.
- Jones, C. N., Wilkinson, K. A., Hung, K. T., Weeks, K. M., & Spremulli, L. L. (2008). Lack of secondary structure characterizes the 5' ends of mammalian mitochondrial mRNAs. *RNA*, 14, 862–871.
- Jørgensen, M. G., Thomason, M. K., Havelund, J., Jørgensen, M. G., Thomason, M. K., Havelund, J., ... Storz, G. (2013). Dual function of the McaS small RNA in controlling biofilm formation dual function of the McaS small RNA in controlling biofilm formation. *Genes & Development*, 27, 1132–1145. <https://doi.org/10.1101/gad.214734.113>
- Julius, C., Riaz-Bradley, A., & Yuzenkova, Y. (2018). RNA capping by mitochondrial and multi-subunit RNA polymerases. *Transcription*, 9, 292–297.
- Julius, C., & Yuzenkova, Y. (2017). Bacterial RNA polymerase caps RNA with various cofactors and cell wall precursors. *Nucleic Acids Research*, 45, 8282–8290.
- Kiledjian, M. (2018). Eukaryotic RNA 5'-end NAD(+) capping and DeNADding. *Trends in Cell Biology*, 28, 454–464.
- Koch-Nolte, F., Fischer, S., Haag, F., & Ziegler, M. (2011). Compartmentation of NAD⁺-dependent signalling. *FEBS Letters*, 585, 1651–1656.
- Konopka, J. B. (2012). N-acetylglucosamine (GlcNAc) functions in cell signaling. *Scientifica*, 2012, 1–15.
- Kowtoniuk, W. E., Shen, Y., Heemstra, J. M., Agarwal, I., & Liu, D. R. (2009). A chemical screen for biological small molecule-RNA conjugates reveals CoA-linked RNA. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 7768–7773.
- Kulbachinskiy, A., & Mustaev, A. (2006). Region 3.2 of the sigma subunit contributes to the binding of the 3'-initiating nucleotide in the RNA polymerase active center and facilitates promoter clearance during initiation. *The Journal of Biological Chemistry*, 281, 18273–18276.
- Kushner, S. R. (2004). mRNA decay in prokaryotes and eukaryotes: Different approaches to a similar problem. *IUBMB Life*, 56, 585–594.
- Liu, S. R., Hu, C. G., & Zhang, J. Z. (2016). Regulatory effects of cotranscriptional RNA structure formation and transitions. *WIREs RNA*, 7, 562–574. <https://doi.org/10.1002/wrna.1350>
- Malecka, E. M., Strozicka, J., Sobanska, D., & Olejniczak, M. (2015). Structure of bacterial regulatory RNAs determines their performance in competition for the chaperone protein Hfq. *Biochemistry*, 54, 1157–1170.
- Malygin, A. G., & Shemyakin, M. F. (1979). Adenosine, NAD and FAD can initiate template-dependent RNA synthesis catalyzed by *Escherichia coli* RNA polymerase. *FEBS Letters*, 102, 51–54.
- Mandin, P., & Gottesman, S. (2009). Regulating the regulator: An RNA decoy acts as an OFF switch for the regulation of an sRNA. *Genes & Development*, 23, 1981–1985.
- Mauer, J., Luo, X., Blanjoie, A., Jiao, X., Grozhik, A. V., Patil, D. P., ... Jaffrey, S. R. (2017). Reversible methylation of m6A in the 5' cap controls mRNA stability. *Nature*, 541, 371–375. <https://doi.org/10.1038/nature21022>
- McLennan, A. G. (2006). The NUDIX hydrolase superfamily. *Cellular and Molecular Life Sciences*, 63, 123–143.
- McLennan, A. G. (2013). Substrate ambiguity among the NUDIX hydrolases: Biologically significant, evolutionary remnant, or both? *Cellular and Molecular Life Sciences*, 70, 373–385.
- Mei, S. C., & Brenner, C. (2014). Quantification of protein copy number in yeast: The NAD⁺ metabolome. *PLoS One*, 9, e106496.
- Mildvan, A. S., Xia, Z., Azurmendi, H. F., Saraswat, V., Legler, P. M., Massiah, M. A., ... Amzel, L. M. (2005). Structures and mechanisms of NUDIX hydrolases. *Archives of Biochemistry and Biophysics*, 433, 129–143.
- Mohanty, B. K., & Kushner, S. R. (2016). Regulation of mRNA decay in bacteria. *Annual Review of Microbiology*, 70, 25–44.
- Moll, I., Grill, S., Gualerzi, C. O., & Blasi, U. (2002). Leaderless mRNAs in bacteria: Surprises in ribosomal recruitment and translational control. *Molecular Microbiology*, 43, 239–246.
- Montange, R. K., & Batey, R. T. (2008). Riboswitches: Emerging themes in RNA structure and function. *Annual Review of Biophysics*, 37, 117–133.
- Negrete, A., & Shiloach, J. (2015). Constitutive expression of the sRNA GadY decreases acetate production and improves *E. coli* growth. *Microbial Cell Factories*, 14, 1–10. <https://doi.org/10.1186/s12934-015-0334-1>
- Nevo-Dinur, K., Nussbaum-Shochat, A., Ben-Yehuda, S., & Amster-Choder, O. (2011). Translation-independent localization of mRNA in *E. coli*. *Science*, 331, 1081–1084.
- Opdyke, J. A., Fofo, E. M., Hemm, M. R., & Storz, G. (2011). RNase III participates in gadY-dependent cleavage of the gadX-gadW mRNA. *Journal of Molecular Biology*, 406, 29–43. <https://doi.org/10.1016/j.jmb.2010.12.009>
- Pulvermacher, S. C., Stauffer, L. T., & Stauffer, G. V. (2009). Role of the sRNA GcvB in regulation of cycA in *Escherichia coli*. *Microbiology*, 155, 106–114. <https://doi.org/10.1099/mic.0.023598-0>
- Ramanathan, A., Robb, G. B., & Chan, S. H. (2016). mRNA capping: Biological functions and applications. *Nucleic Acids Research*, 44, 7511–7526.
- Reichenbach, B., Maes, A., Kalamorz, F., Hajnsdorf, E., & Gorke, B. (2008). The small RNA GlmY acts upstream of the sRNA GlmZ in the activation of glmS expression and is subject to regulation by polyadenylation in *Escherichia coli*. *Nucleic Acids Research*, 36, 2570–2580. <https://doi.org/10.1093/nar/gkn091>
- Ross, W., Sanchez-Vazquez, P., Chen, A. Y., Lee, J. H., Burgos, H. L., & Gourse, R. L. (2016). ppGpp binding to a site at the RNAP-DksA Interface accounts for its dramatic effects on transcription initiation during the stringent response. *Molecular Cell*, 62, 811–823.
- Rutherford, S. T., Lemke, J. J., Vrentas, C. E., Gaal, T., Ross, W., & Gourse, R. L. (2007). Effects of DksA, GreA, and GreB on transcription initiation: Insights into the mechanisms of factors that bind in the secondary channel of RNA polymerase. *Journal of Molecular Biology*, 366, 1243–1257.
- Sarkar, N. (1996). Polyadenylation of mRNA in bacteria. *Microbiology*, 142(Pt 11), 3125–3133.
- Shell, S. S., Wang, J., Lapierre, P., Mir, M., Chase, M. R., Pyle, M. M., ... Gray, T. A. (2015). Leaderless transcripts and small proteins are common features of the mycobacterial translational landscape. *PLoS Genetics*, 11, e1005641.

- Silva, I. J., Saramago, M., Dressaire, C., Domingues, S., Viegas, S. C., & Arraiano, C. M. (2011). Importance and key events of prokaryotic RNA decay: The ultimate fate of an RNA molecule. *WIREs RNA*, 2, 818–836.
- Stein, L. R., & Imai, S. (2012). The dynamic regulation of NAD metabolism in mitochondria. *Trends in Endocrinology and Metabolism: TEM*, 23, 420–428.
- Tamm, J., & Polisky, B. (1985). Characterization of the ColE1 primer-RNA1 complex: Analysis of a domain of ColE1 RNA1 necessary for its interaction with primer RNA. *Proceedings of the National Academy of Sciences of the United States of America*, 82, 2257–2261.
- Topisirovic, I., Svitkin, Y. V., Sonenberg, N., & Shatkin, A. J. (2011). Cap and cap-binding proteins in the control of gene expression. *WIREs RNA*, 2, 277–298.
- Vasilyev, N., & Serganov, A. (2015). Structures of RNA complexes with the *Escherichia coli* RNA pyrophosphohydrolase RppH unveil the basis for specific 5'-end-dependent mRNA decay. *The Journal of Biological Chemistry*, 290, 9487–9499.
- Vogel, U., & Jensen, K. F. (1994). The RNA chain elongation rate in *Escherichia coli* depends on the growth rate. *Journal of Bacteriology*, 176, 2807–2813.
- Vvedenskaya, I. O., Bird, J. G., Zhang, Y., Zhang, Y., Jiao, X., Barvik, I., ... Nickels, B. E. (2018). CapZyme-Seq comprehensively defines promoter-sequence determinants for RNA 5' capping with NAD. *Molecular Cell*, 70, 553–564.e9.
- Wagner, E. G. H. (2013). Cycling of RNAs on Hfq. *RNA Biology*, 10, 619–626. <https://doi.org/10.4161/rna.24044>
- Walters, R. W., Matheny, T., Mizoue, L. S., Rao, B. S., Muhlrud, D., & Parker, R. (2017). Identification of NAD⁺ capped mRNAs in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America*, 114, 480–485.
- Waters, L. S., & Storz, G. (2009). Regulatory RNAs in bacteria. *Cell*, 136, 615–628.
- White, H. B., 3rd. (1976). Coenzymes as fossils of an earlier metabolic state. *Journal of Molecular Evolution*, 7, 101–104.
- Wimpenny, J. W., & Firth, A. (1972). Levels of nicotinamide adenine dinucleotide and reduced nicotinamide adenine dinucleotide in facultative bacteria and the effect of oxygen. *Journal of Bacteriology*, 111, 24–32.
- Wu, P., Liu, X., Yang, L., Sun, Y., Gong, Q., Wu, J., & Shi, Y. (2017). The important conformational plasticity of DsrA sRNA for adapting multiple target regulation. *Nucleic Acids Research*, 45, 9625–9639. <https://doi.org/10.1093/nar/gkx570>
- Yao, S., Sharp, J. S., & Bechhofer, D. H. (2009). *Bacillus subtilis* RNase J1 endonuclease and 5' exonuclease activities in the turnover of DeltaermC mRNA. *RNA*, 15, 2331–2339.
- Zhang, D., Liu, Y., Wang, Q., Guan, Z., Wang, J., Liu, J., ... Yin, P. (2016). Structural basis of prokaryotic NAD-RNA decapping by NudC. *Cell Research*, 26, 1062–1066.

How to cite this article: Julius C, Yuzenkova Y. Noncanonical RNA-capping: Discovery, mechanism, and physiological role debate. *WIREs RNA*. 2019;10:e1512. <https://doi.org/10.1002/wrna.1512>

Review Article

Bacteriophage gene products as potential antimicrobials against tuberculosis

Maria Puiu and  Christina Julius

Centre for Bacterial Cell Biology, Newcastle University, Baddiley Clark Building, Richardson Road, Newcastle Upon Tyne NE2 4AX, U.K.

Correspondence: Christina Julius (C.Julius2@ncl.ac.uk)

Tuberculosis (TB) is recognised as one of the most pressing global health threats among infectious diseases. Bacteriophages are adapted for killing of their host, and they were exploited in antibacterial therapy already before the discovery of antibiotics. Antibiotics as broadly active drugs overshadowed phage therapy for a long time. However, owing to the rapid spread of antibiotic resistance and the increasing complexity of treatment of drug-resistant TB, mycobacteriophages are being studied for their antimicrobial potential. Besides phage therapy, which is the administration of live phages to infected patients, the development of drugs of phage origin is gaining interest. This path of medical research might provide us with a new pool of previously undiscovered inhibition mechanisms and molecular interactions which are also of interest in basic research of cellular processes, such as transcription. The current state of research on mycobacteriophage-derived anti-TB treatment is reviewed in comparison with inhibitors from other phages, and with focus on transcription as the host target process.

Introduction

Mycobacterium tuberculosis, the infective agent of tuberculosis (TB), is one of the deadliest human pathogens, with 1.7 million related deaths reported in 2016. Tuberculosis infection in otherwise healthy individuals can remain undetected for a long time. Thus, the disease is expected to be far more common than documented. Twenty-three percent of the world population is estimated to be carrier of latent TB [1], 10 million new cases of active TB were reported in 2017 globally [1]. Active TB can affect all body tissues, though it most commonly manifests in the respiratory system. Death results from extensive tissue damage which can be accompanied by hypoxia (pulmonary TB) or seizures (TB meningitis) [2]. It is challenging to treat *M. tuberculosis*, as well as other infections caused by *Mycobacterium*, such as Buruli ulcer (caused by *Mycobacterium ulcerans*), due to the unique physiology of the organism. For a review on *M. tuberculosis* virulence, see [2]. Mycobacteria have an unusually high lipid content cell wall which is impermeable to many drugs and hydrophilic molecules [3]. Furthermore, it is an intracellular pathogen; therefore, the drug not only has to penetrate the bacterial cell envelope but also the host cell (macrophages, dendritic cells or tissue cells, e.g. alveolar cells). Above all, *M. tuberculosis* grows very slowly and after the initial infection can remain dormant for months, and non-growing cells make a poor target for many drugs. Infection foci can turn into granuloma that are formed by the immune system to contain the pathogen and are exploited by the bacterium to remain dormant until the immune system is weakened. Then the granuloma, which are otherwise able to contain the infection indefinitely, are liquefied and bacterial growth resumes [2].

WHO has named research and innovation as the 'third pillar' of the End TB Strategy. Besides increased efforts in improvement of and addition to the antibiotic repertoire of current treatment, research on TB vaccines is intensified. However, the currently stated goal to reach complete eradication of TB by 2035 requires a research breakthrough both in TB diagnostics and treatment to which mycobacteriophage research should be considered as a potential contribution.

Received: 5 November 2018
Revised: 2 April 2019
Accepted: 8 April 2019

Version of Record published:
13 May 2019

Current tuberculosis treatment and treatment alternatives: more is more

The current first-line treatment of drug-sensitive TB (DS-TB) involves a 6-month course of rifampicin (inhibition of bacterial RNAP), isoniazid (inhibition of fatty acid synthase with effect on cell wall), ethambutol (inhibition of arabinosyl transferase with effect on cell wall) and pyrazinamide (inhibition of cell wall synthesis via fatty acid synthase) [4]. The combination drug treatment was developed in 1960 and the treatment duration was reduced from 18 to now 6 months [5]. Rifampicin revolutionised treatment as it acts on dormant cells [6]. The drug cocktail reduces the risk of resistance development, which still can arise especially if treatment is paused or prematurely aborted [7]. Drug-resistant (DR) infections are documented increasingly often [8]. One common cause of first-line drug resistance is resistance to rifampicin via mutation of RNA polymerase (RNAP) at the rifampicin-binding pocket [6]. Strains resistant to at least rifampicin and isoniazid are considered multidrug-resistant (MDR). Drugs which are currently used in clinical trials, such as bedaquiline (inhibition of ATP synthase), delamanid (cell wall synthesis inhibitor) or linezolid (blocks ribosome activity), classify as third-line treatment as they are used in anti-MDR-TB treatment, but clinical trials are incomplete [5]. If the resistance profile also includes second- or third-line medications, this is described as extensively drug-resistant (XDR) [5]. Pan-resistant strains have been documented [9,10].

Treatment of drug-susceptible (DS)-TB is successful in 86% cases, MDR-TB is cleared in only 50% cases. MDR/XDR-TB therapy can constitute up to 2 years [11,12]. Long-term antibacterial treatment has detrimental effects on the natural microflora and thus promotes further health complications in addition to common side effects of the drugs themselves. Side effects can result in abortion of treatment thus the infection is prolonged, and resistance is more likely to emerge. Second- and third-line drugs prove to have higher toxicity to the patient, and higher cost to the health care system [5]. Furthermore, the chances to transmit the disease to others during prolonged treatment are higher. WHO thus sees an update in the treatment regimen as desirable [1].

Alternatives to antibiotic treatment are commonly used in economically poor or rural areas where medicine is not available. In provinces in South Africa, TB is often treated with medicinal herbs. The choice of an herb (e.g. *Artemisia afra*, *Carica papaya* or *Myrothamnus flabellifolius*) is according to local traditions and beliefs but their effectivity against TB is documented in the literature [5]. Such plant-based treatments could realistically constitute an addition to the antibiotic regimen, though in areas where prescription drugs are available should not replace them. Research on plant-derived drugs could potentially be based on those traditionally used herbs in the future.

Tuberculosis vaccine consisting of an attenuated strain of *Mycobacterium bovis*, the Bacille Calmette–Guerin (BCG) vaccine, has been in use since 1921. It is not sufficiently effective to be a reliable preventive measure [13], as it does not prevent primary or reactivated infection but helps cure acute and disseminated disease [5]. Its effectiveness varies across geographical location and ethnic group and also in dependence on the endemic TB strain — and was documented to be between 0% and 83% [13]. Reasons for conflicting numbers in different studies are likely molecular differences between BCG strain and TB strain, age and genetic background of the patient, climate, vaccination dose, and other potentially unknown factors [13]. Vaccination of immunocompromised persons is not recommended as it can lead to complications [13]. Still, BCG vaccine is routinely administered to newborns of Aboriginal Canadians and people at high risk to contract MDR-TB (e.g. health care workers in dense MDR-TB areas) [13]. Variants of BCG vaccine boosters are currently being tested. Those may contain fragmented *M. tuberculosis* as an antigen, e.g. RUTI, or recombinant fusion proteins, e.g. AERAS-402 which is a replication-deficient adenovirus that expresses virus-*M. tuberculosis* fusion proteins. Vaccines under study to replace the inefficient BCG vaccine are based on *M. vaccae* or recombinant *M. bovis* strain VMP2002 [5].

Another treatment method for TB is immunotherapy. Several immune therapies are in use or in a clinical trial to be used as an adjunct to combination drug therapy to obtain a reduction in symptom severity and duration. Application of cytokines and immune modulators are being tested. For example, the subcutaneous application of interferon- γ results in reduced inflammation of the infected lung [14]. Interestingly, administration of hepatitis C virus and hepatitis B virus antigens (called Immunitor-V5) has been shown to aid effectively in TB clearance while *M. vaccae* antigen (also called V7) studies are conflicting [5,15–17]. For more information on experimental TB drugs, see [5].

Mycobacteriophages—potential and challenges of phage therapy against tuberculosis

All TB treatment options will need to overcome the challenges the infection proposes (tissue/granuloma penetration, host cell penetration, drug interaction with HIV therapy). Additionally, they have to be low in toxicity and adverse effects on microflora, short in duration and will have to be made available in rural and poor areas. Therapy with live mycobacteriophages may comply if the research can solve a few problems which shall be described further in this chapter. The main advantages of phage therapy are low cost of production, sparing of microflora, and autoregulation of phage levels in the patient. Negative interactions between phage and chemical drugs have not been documented.

Genomic studies have revealed that mycobacteriophages are one of the most diverse groups of bacteriophages so far characterised. Mycobacteriophage genomes present highly mosaic structures due to horizontal gene transfer and history of host changes. Over 1720 mycobacteriophage genomes have been sequenced to date [18] and arranged into clusters according to shared nucleotide sequence similarities. As Figure 1 illustrates, there are a total of 29 clusters, which share little or no sequence similarity, and several singletons (phages with no close relatives). The clustered genomes share features such as regulatory systems or transfer RNA sequences [19]. Unique variety in the mycobacteriophages makes them fascinating study objects since they offer insight into mechanisms of phage evolution and diversity [20]. The potential to find new regulatory mechanisms of the mycobacterium–phage interaction is therefore high.

Mycobacteriophages, like all bacteriophages, may execute one of two types of infection: lytic or lysogenic. The lytic cycle is adopted by virulent phages and is initiated after adsorption of the phage to host cell receptor and injection of phage DNA into the cytoplasm. Upon transcription of viral genes and replication of its genome, phage structural proteins are produced and organised into capsids (or ‘heads’) and tails. With the association of daughter genomes and capsid proteins, the virion is formed. At the end of the lytic cycle, the host cell wall is enzymatically lysed. This leads to the release of newly formed viral particles which can undertake the same infectious cycle [21]. Temperate phages, which may opt for a lysogenic life cycle, do so by integrating into the host chromosome using integrases and *attP* and *attB* sites for homologous recombination. They remain dormant by selectively expressing only the phage lytic repressor protein and phage DNA is

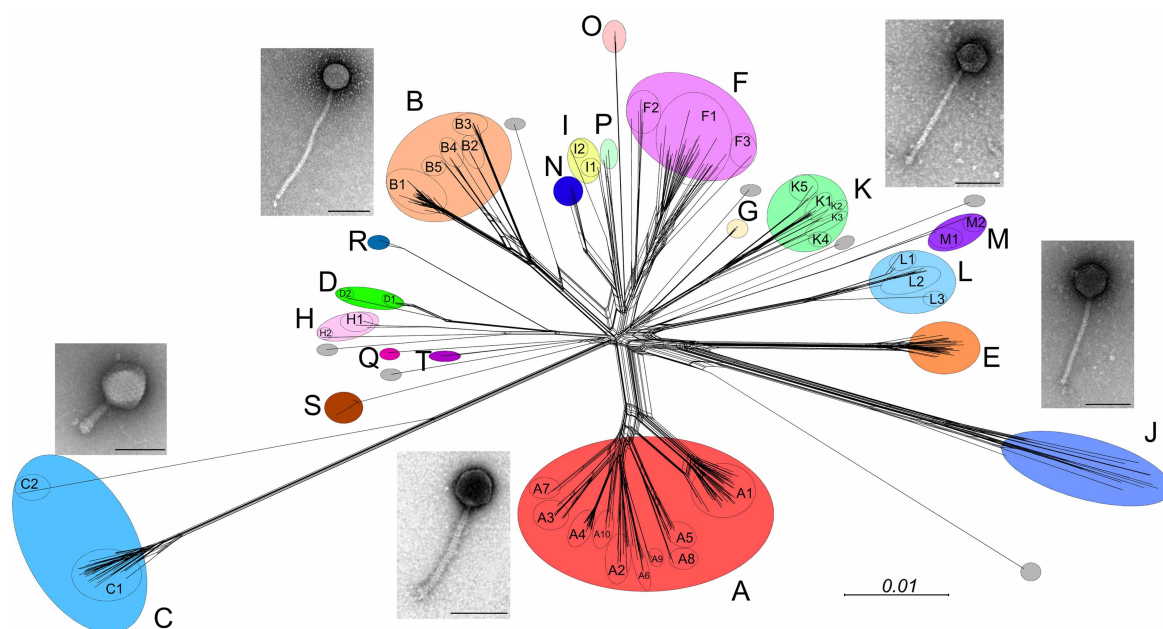


Figure 1. Phylogenetic tree displaying sequenced genomes of 471 mycobacteriophages as analysed by Hatfull [19].

The tree captures the diversity of mycobacteriophages, which also represent the largest class of phages with sequenced genomes to infect a single bacterial strain, *Mycobacterium smegmatis* mc2155 [19]. The genomes have been classified into 29 clusters (clusters A, B, C, etc.) and further in subclusters (e.g. B1, B2) according to their shared sequence similarities.

replicated alongside the host chromosome by host machinery. A lytic cycle can be triggered by different stress conditions, predominantly those that lead to DNA damage response via RecA pathway [22,23]. The repressor is deactivated and lytic and structural genes can be expressed in analogy to lytic phage gene expression. Some phages have a pseudotemperate lifestyle in which their DNA does not integrate within the host chromosome, but lytic growth is still repressed and the phage genome replicates in the form of a plasmid [24].

Treatment of TB with live phages has been demonstrated by Sula et al. [25] in an animal model. They reported decreased lesions in the lungs, liver and spleen and reduction in granuloma formation when testing mycobacteriophage DS6A in *M. tuberculosis*-infected guinea pigs [26]. Another study documents successful treatment of Buruli ulcer (*M. ulcerans*) on mouse feet using mycobacteriophage D29 subcutaneous injection [27]. In theory, a live phage is its own delivery system. If phage is administered to the site of infection (skin, airways, bloodstream, etc.), it is able to recognise its host, infect, kill and move on to the next host. By multiplying inside the host cells, phage concentration at the site of infection will automatically increase, and phage titre will cease after eradication of bacterium. This has been shown to be performed by a variety of other phages ([28] and references therein). For example, phage therapy was successful in a *Klebsiella pneumoniae* lung infection [29] and against *Staphylococcus aureus* septicaemia and abscesses on organs in mice [30]. Both *K. pneumoniae* and *S. aureus* can adopt an intracellular lifestyle within macrophages and phages successfully cleared intracellular infection of macrophages *in vitro* [29,30]. Also, *M. tuberculosis* is an intracellular pathogen, and thus far delivery of mycobacteriophages on their own across the macrophage cell membrane has not been achieved [31]. Broxmeyer et al. used non-pathogenic *M. smegmatis* cells to transport TM4 phage particles to *M. tuberculosis* in macrophages. TM4 is a lytic mycobacteriophage which is specific to both hosts, *M. tuberculosis* and *M. smegmatis* [32]. The technique led to a substantial decrease in TB titre, suggesting that TM4 successfully infected and killed *M. tuberculosis*. A further issue for mycobacteriophage delivery is granuloma formation [28]. Granuloma solid and layered structure is likely to interfere with phages reaching their host [33]. This problem has not yet been tackled in mycobacteriophage therapy research.

Opponents of phage therapy state eradication of phage by the human immune system would abort treatment and prevent consecutive treatments. Even though phages do not attack human cells, presence of foreign material in the body will lead to immunoglobulin-directed phage inactivation, phagocytosis and immune memory against the phage [34–36]. Special delivery systems that protect phage from antibodies need to be constructed to overcome this problem. Chhibber and co-workers reported the use of liposomes to shield phage from antibody in anti-phage serum from mice [35]. *In vitro*, phage without liposome was inactivated within 3 h in blood serum while liposomal protection was 100% efficient (no inactivation, no reduction in titre). Liposomes did not interfere with phages' ability to adsorb to bacterial cells. The same study also showed phage delivery into macrophages was not hindered by liposomes to clear intracellular *K. pneumoniae* infection. Thus, this method is potentially viable in TB therapy as well.

Other studies even observe a synergy between phage and immune system, coined 'immunophage synergy' [37,38]. In a mouse model of *P. aeruginosa* lung infection, the contrast between the phage elimination from the tissue by the animal immune system, and the self-amplification of the phage at the site of infection, allowed for phage population to remain stable, furthering pathogen lysis [38]. Also, phage–antibiotic synergy (PAS) has been proposed and documented in several studies [39–42] with different pathogens. Phage characteristics such as good biofilm penetration [42] and expression of lytic enzymes [39] potentiate the drug. Phage-induced holins in the bacterial cell envelope support activity of cell wall destabilising antibiotics or reactive oxygen species [39] to an extent that can reverse antibiotic resistance [40]. This way, also alternative treatments such as honey were enhanced [43].

Antibacterial effectors of phage origin: enzybiotics as potential drugs

Acceptance of phage therapy is low in Western medicine. Reasons for this are the necessity to determine the infective agent prior start of therapy due to host specificity of the phage (which in the case of TB is not an issue); interaction with (and possibly clearance by) the human immune system; necessity to eliminate sterilisation during drug manufacturing (virus elimination as part of regulated drug production); and limitations in phage administration due to nature of phage or disease, as described above for TB in granuloma. However, phages represent a rich source of potential new antibacterials, with new mechanisms of action which are not

affected by pre-existing antibiotic resistance and could potentially be implemented in the treatment regimen of MDR/XDR, as well as DS-TB [44–46].

The evolutionary arms race between phages and their hosts has previously been an exciting source for molecular tools. Bacteriophages have evolved alongside their bacterial hosts, which consequently developed numerous defence mechanisms such as the CRISPR/Cas system or restriction endonucleases [20,47]. Phages have, in turn, adapted to host bacterial immunity by evolving mechanisms of defence from spontaneous mutations [48]. Additionally, anti-CRISPR proteins can be found on phage genomes (which inhibit Cas protein activity or cRNA transcription or maturation) [47]. Mycobacteriophage genomes are well characterised. 1720 genomes have been sequenced, with ORFs identified, yielding a large database of putative protein-coding sequences. Among the non-structural genes, although, the proteins encoded are not well characterised, many were shown to be cytotoxic in the mycobacterial host. A number of these also reduce host gene expression [49–51], also called host shut-off, which is a typical feature of viral infection [49]. These toxic phage proteins are good candidates for future antibacterials.

A few potential mycobacteriophage effectors are currently under investigation. It was proposed that phage endolysins can be administered exogenously and that such antibacterial enzymes should be called enzybiotics [52,53]. At the end of the lytic life cycle, a phage must lyse the bacterial cell wall in order for its progeny to be set free into the medium. This is accomplished by a two-component system; a holin that creates holes in the inner cell membrane and an endolysin that is enriched in the cytosol during phage growth and then leaks through the pores. The endolysin is a peptidoglycan hydrolase; it digests the cell wall. Host specificity is one of the advantages of phage gene products over antibiotics. Phages infect and kill via a cascade of molecular interactions which are highly adapted to the host target structures, from cell surface receptors to epitopes on target host proteins. Also, the phage peptidoglycan hydrolases show high specificity to a type of peptidoglycan of one genus or even serotype [52–54]. Mycobacteriophage Ms6 genome encodes an endolysin LysA and a lipolytic enzyme LysB [55]. LysA was previously described as an amidase which cleaves the bond between two essential amino acids that make up the glycosidic bond, which is crucial in the peptidoglycan formation in bacteria [55]. LysB has a high affinity for longer chain length substrates and brings additional lipolytic activity to hydrolyse the very complex mycobacterial cell wall [56]. Restricted access of peptidoglycanases to cell wall through the outer membrane has been proposed to complicate administration of endolysins from the outside [57]. Though cell division might allow to overcome the outer membrane, such that growing cells might indeed be susceptible to endolysins “from without” [58]. Some antibacterial activity was shown for *M. smegmatis* intracellular infection in macrophages. Both lysins of mycobacteriophage BTCU-1, LysA and LysB, induced morphological changes to *M. smegmatis* cultures and reduced titre of *M. smegmatis* intracellular infections [59]. Mycobacteriophage lysis proteins could also be of benefit as supporting treatment to antibiotics, similar to the previously described synergy between phage and antibiotic. For example, mycobacteriophage Bxb1 is a lytic phage, which forms plaques with halos. Plaque centres are composed of lysed cells. In the case of other phages, halos are often caused by an enzymatic component diffusing through top agar and inducing changes to the bacterial cell wall, such as glucanases [60]. Such enzymes might be studied for their potential to aid other therapies in the future.

Also in other phage–host systems, phage proteins are under investigation for antibacterial application. Wang et al. [61] engineered a fusion protein to aid uptake of phage endolysin into mammalian cells. *Staphylococcus aureus* phage JD endolysin JD_{lys} was fused to a cell penetrating transcription activator CPP_{tat} (to give CPP-JD_{lys}). CPP-JD_{lys} was successfully used to treat murine intracellular skin infection with Methicillin-resistant *S. aureus* (MRSA), a problematic nosocomial pathogen [61]. Such a protein-based delivery system for phage effectors could be applicable for mycobacteriophage lysis or other cytotoxic mycobacteriophage proteins as well. Mycobacteriophage enzybiotics could also be employed in infection prevention, as has been reported for *Streptococcus* infections: Application of recombinant phage endolysin prevented bacterial infection in the mouse model [62]. Enzybiotics can be taken even further in transgenic livestock: Wall showed prevention of mastitis in transgenic cows expressing *S. aureus* phage endolysin [63].

Large-scale screening procedures may help to identify new anti-TB drugs. Wei et al. synthesised 200 peptides based on mycobacteriophage gene databases and screened them for their activity on trehalose-6,6'-dimycolate, an *M. tuberculosis* glycolipid. They found one protein, PK34, which binds the glycolipid and cleared TB in the mouse model, and reduced level of proinflammatory cytokines [64]. Further studies should be undertaken which might focus on a specific mycobacterial target and use toxicity screening methods or pull-down assays to efficiently find mycobacteriophage effectors. For a review on potential anti-TB peptides from various sources, see [65].

Host transcription as a target of bacteriophage regulators

One potential mycobacterial target is transcription, which is the first and most regulated step of gene expression. The main enzyme of transcription, RNAP makes a good target for inhibition because of its central role in the cell and its conserved nature in the prokaryotic kingdom which allows for selective targeting. The basal level of transcription is needed even in the dormant stage [66], therefore, non-growing cells can be targeted with drugs against RNAP. The process of transcription can be segregated into several steps (see Figure 2). These are (i) holoenzyme formation and promoter binding, (ii) melting of double-stranded DNA around the transcription start site, and isomerization from RNAP-promoter closed complex (R_{PC}) to

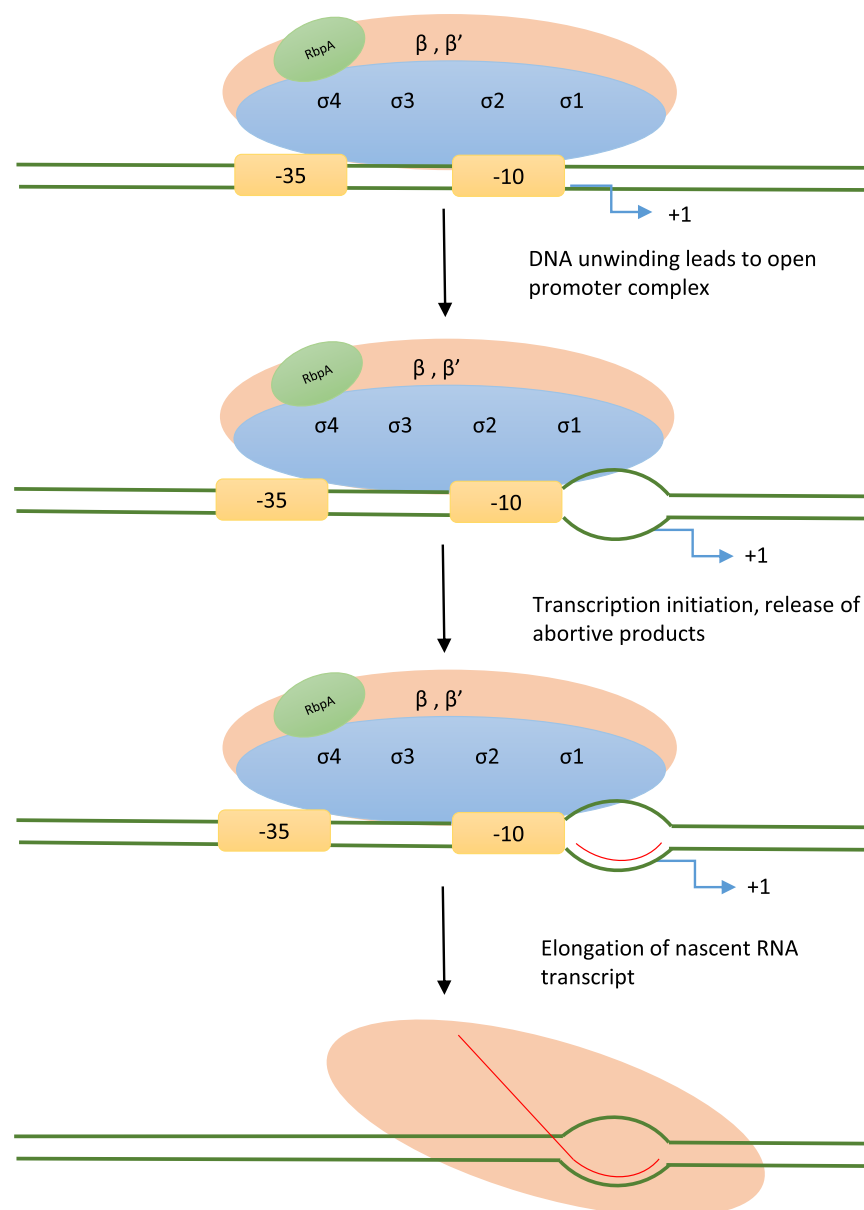


Figure 2. Schematic overview of transcription cycle in bacteria.

RNAP β and β' subunit are depicted in pink, and the different domains of σ subunit (σ_1 , σ_2 , σ_3 , σ_4) are depicted in blue. Upon binding of the holoenzyme to the promoter (σ_4 to the -35 region, and σ_2 to the -10), DNA unwinds, leading to an open complex. This is followed by transcription initiation (signalled as the +1 arrow) and release of short RNA abortive products. Elongation stage is defined by the production of RNA chains ≥ 10 nucleotides long [62].

RNAP-promoter open complex (RPO), (iii) initiation of RNA synthesis, (iv) promoter escape and elongation and finally (v) transcription termination (not shown in the figure). The RNAP core is made up of several subunits, α , β , β' , γ and ω (in Figure 2 only β , β' are shown for simplicity), completed by the σ factor to form the holoenzyme. Essential transcription factor RbpA is a common feature in Actinobacteria (such as *Mycobacterium*) RNAP-promoter complexes. RbpA is a protein that binds the RNAP holoenzyme and has a role in transcription initiation. It was hypothesised to play a role in promoter binding and open complex formation on non-extended $-10'$ -type promoters as well as σB promoters [67,68]. The elongation stage of transcription takes place after dissociation of core enzyme from σ and promoter region in progressive fashion along the template in open complex. Continuous RNA polymerisation can be interrupted by pausing and termination signals but can be resumed by the transcription elongation complex (TEC). Errors in transcription (e.g. mismatches) can lead to backtracking of RNAP [69]. Transcription is terminated when RNAP reaches a termination signal. Rho-dependent and Rho-independent termination function via two different mechanisms. In Rho-independent termination, secondary structure formation in the RNA product (loops) destabilise TEC and cause RNAP to pause and eventually dissociate from the template. In *Mycobacteria* this depends on 7–9 nt long U-rich sequences (U-tracts). *Mycobacterial* RNAP (in contrast with *E. coli* RNAP) terminates efficiently at U-tracts even in the presence of elongation factors. Furthermore, the elongation factor NusG in mycobacterial transcription can perform a second terminatory role [70]. Rho-dependent termination is characterised by destabilisation of the TEC over termination factor Rho [71,72]. Similar to the *E. coli* Rho-dependent pathway, mycobacterial Rho binds to sites on mRNA and uses ATPase and helicase activity to move itself towards RNAP to initiate its release from the template. Transcription of ribosomal genes is terminated in a different mechanism involving Nus proteins and other factors [72]. RNAP gene sequences are vastly conserved throughout the prokaryotes and altogether ~25% of sequences show evolutionary variability. Variable regions are mostly expendable and their deletion does not affect basic enzymatic activity. However, they provide targets for species-specific regulators as well as exogenous interaction partners, like phage-encoded regulators [73].

Thus far, no phage RNAP has been found to be encoded by any known mycobacteriophage. In lysogenic mycobacteriophage L5, it was shown that phage gene expression is susceptible to rifampicin, indicating that the phage requires bacterial RNAP [50]. Thus, it is very likely that mycobacteriophages regulate host RNAP, as many other known phages do [74]. It was also shown that L5 lytic gene expression results in host shut-off [50]. However, for mycobacteriophage TM4, no host shut-off could be observed during lytic infection [75]. Figure 3 depicts how mycobacteriophage gene expression is regulated in the context of the phage life cycle. Most of what is known about mycobacteriophage transcription regulation concern transcription patterns over the course of an infection. In lytic infection, gene expression is often regulated to take place in phases [75]. For example, TM4 gene expression is regulated to take place in an early phase (20–60 min p.i. [post infection]) and a late phase (60 min p.i. and later) [75], thus co-ordinated transcription regulation must take place during TM4 infection as well. The late genes seem to be mainly associated with virus particle structure and assembly [75]. TM4 also encodes a protein with high similarity to *Streptomyces* transcriptional regulator WhiB, a transcription factor of sporulation [75]. *M. tuberculosis* also carries seven *whiB*-like genes (*whiB1-whiB7*). Overexpression of TM4-*whiB* inhibits mycobacterial cell division via specific repression of *whiB2* (a component of cell division regulation) and is thus a putative candidate for anti-TB therapy research [76]. Fan et al. [77] were the first to investigate the reciprocal effects of mycobacteriophage gene expression (phage SWU1) and host gene expression (*M. smegmatis*). SWU1 infection shifted host gene expression rather than causing a shut-off. While genes of RNA translation and degradation, protein export and other secretion and glycerophospholipid metabolism were up-regulated, expression of siderophores and nitrotoluene degradation genes were down-regulated. Markedly, SWU1 infection up-regulates translation of RNAP subunits genes *rpoA*, *rpoB*, *rpoC* and *rpoZ* as well as housekeeping σA and stress sigma factor σF [77]. Up-regulation of transcription machinery combined with strong promoters is a reasonable method of viruses to ensure their own transcription without suppressing host maintenance. *Mycobacteria* have a variety of σ factors (*M. tuberculosis* has 13, *M. smegmatis* 28) with σA being the housekeeping σ [78]. σA is also the only $-10/-35$ type σ with sequence specificity in analogy to *E. coli* $\sigma 70$. Mycobacteriophage BP, like many other mycobacteriophages, potentially usurps transcription with a strong promoter that interacts with host σA [78].

Another aspect of transcription known to be manipulated by phages is termination. Many phages have long operons with many open reading frames (ORFs). To ensure transcription of long RNA, termination signals

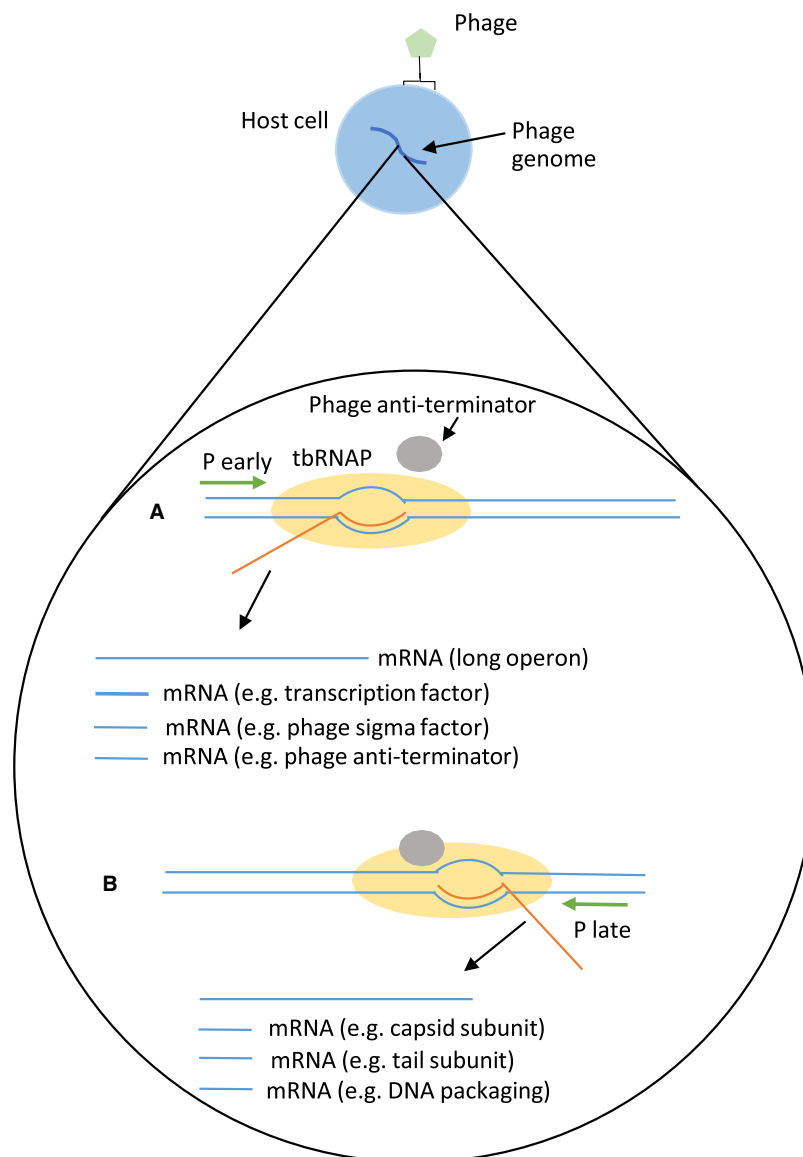


Figure 3. Illustration of phage transcription in the context of phage life cycle.

Upon adsorption onto the bacterial cell, the phage injects its genome into the host. Temperate phages integrate their genome into the host chromosome to reside dormant (as a lysogen) and express a lytic repressor which can be suppressed to induce the lytic life cycle. In case of lytic phages, phage gene expression follows genome injection [21,23]. **(A)** Host RNAP (tbRNAP, yellow) recognises the strong phage promoter (here P early) to transcribe mainly non-structural genes used to regulate cellular and phage processes such as for transcription factors, anti-terminators or nucleases for mRNA maturation. Anti-termination systems (here in grey) may be used to ensure long phage operons are entirely transcribed. **(B)** Later in infection, transcription of late genes is started (here from another promoter P late) which usually encode structure and assembly proteins such as capsid or tail subunits, proteases for capsid maturation or DNA packaging proteins. Finally, the enrichment of continually expressed lysins lead to cell rupture and mature virions will be released [23].

have to be ignored by RNAP [79,80]. Mycobacteriophage Ms6 was first shown to employ anti-termination by a phage-encoded antiterminator factor [81]. A recent study [82] has demonstrated that a capsid protein (Psu) of *E. coli* host bacteriophage P4 suppresses transcription terminator factor Rho by inhibiting its ATPase and translocase activity, therefore, promoting anti-termination. Psu also inhibits Rho derived from other hosts, including *M. smegmatis*, *M. tuberculosis* and *M. bovis*.

Table 1. Known inhibitors of RNAP or transcription processes from phage gene products or (developmental) drugs such as antibiotics
Published compound name/phage protein are given as referenced in literature, slash means compound or function was referred to in additional names.

Target	Phage effector		Other inhibitory substances	
	Protein (phage, host)	Function (reference)	Inhibitor (classification)	Function (reference)
Initiation				
Sigma binding	AsiA (T4, <i>E. coli</i>)	Anti- σ appropriation [80–82]	Compound C5 (developmental drug)	Blocks β' σ binding site [76]
	G1ORF67/Gp67 (G1, <i>S. aureus</i>)	Anti- σ [83]		
	Gp39 (P23-45, <i>T. thermophilus</i>)	σ -appropriation [32,84]		
Promoter binding	AsiA + MotA (T4, <i>E. coli</i>)	Redirection of RNAP to phage promoter [85–87]		
	G1ORF67/Gp67 (G1, <i>S. aureus</i>)	Redirection of RNAP from UP-element host promoters to phage promoters [83,88]		
Open complex formation	Gp2 (T7, <i>E. coli</i>)	Destabilises RPo [89,90]	Fidaxomicin (antibiotic)	Destabilises RPo via σ and RbpA [91]
	Gp76 (P23-45, <i>T. thermophilus</i>)	Redirection of RNAP from host –35/–10 to phage extended –10 promoters [84,92])	Myxopyronin (antibiotic candidate)	Blockage of switch region [91]
Initiation catalysis			Rifampicin (antibiotic)	Blockage of RNA polymerisation >3 nt [93,94]
			Kanglemycin (antibiotic candidate)	Interferes with phosphodiester bond formation [79]
			GE23077 (antibiotic candidate)	Binds active centre, prevents NTP binding [95]
Elongation				
Ternary elongation complex (TEC) catalysis			Salinamide (antibiotic candidate)	Allosteric alteration of active site [96]
			Streptolydigin	Stabilises inactive intermediate of TEC [97]
Pausing	P7 (Xp10, <i>X. oryzae</i>)	Stabilisation of TEC to inhibit pausing [98]	Microcin J25 (peptide antibiotic)	Blocks secondary channel [99]
Termination				
Termination	N-protein (λ , <i>E. coli</i>)	Binding of elongation factors creates termination-resistant TEC [100,101]		
	p7 (Xp10, <i>X. oryzae</i>) Gp39 (P23-45, <i>T. thermophilus</i>)	Binding of β' stabilises TEC [102] Stabilisation of TEC to transcribe through poly (U) termination signal [103]		

The state of research on bacteriophage effectors on transcription has also advanced in other phage–host systems such as *E. coli*, *Bacillus* or *Thermophilus* phages. Examples of regulators of transcription from phage origin along with known antibiotics targeting the same stage of transcription are given in Table 1. Remarkably, phage effectors' and antibiotics' molecular targets often differ, supporting the hypothesis that phage enzybiotics should remain active against DR pathogens.

Another necessary research focus should be drug delivery. The intracellular nature of *M. tuberculosis* means that second barrier, the mammalian cell membrane, has to be crossed by the drug. In phage therapy, one strategy to overcome the issue could be to feed phage-infected *M. smegmatis* to macrophages [32]. However, macrophages are specialised bacteria eaters. If tissue cells are infected, the eukaryotic membrane needs to be crossed via regular phagocytosis/endocytosis. In enzybiotics, the development of fusion proteins that are able to

transgress mammalian membrane [61], is a good example of where future research might lead us. Finally, in TB, drug delivery is further complicated by granuloma formation. A strategy to overcome this dense tissue has not been found yet.

Conclusion

The study of mycobacteriophage effectors is currently in its infancy. As research of other phage–host systems shows, however, the chances are good to find a mycobacteriophage protein that inhibits host transcription irrespective of antibiotic resistance. Bacteriophage effectors have been found which interact with RNAP, promoter DNA, or other transcriptional regulators. Due to the unprecedented diversity of mycobacteriophages and advanced state of documentation of mycobacteriophage genomes, focussed research for anti-TB drugs from phage origin could be a potential solution to spreading antibiotic resistance of the pathogen. A major problem of TB treatment is drug delivery, due to unique cell envelope and lifestyle of the pathogen. The field of enzybiotics works on protein-based drug design: Fusion proteins of a cell wall translocator and a phage effector protein have been achieved in the context of MRSA skin infections and could be attempted for TB [104–109].

Perspectives

- Importance: TB is one of the leading causes of mortality among bacterial pathogens worldwide, with an estimated 1.3 million deaths occurring every year. WHO has implemented a strategy to eventually eliminate the TB pandemic until 2035. Progress of the strategy is monitored annually and compared with intermediate goals set for 5-year periods. Even though the global burden of the infection has been reduced, progress remains behind set intermediate goals, and there is an increasing emergence of MDR and XDR-TB cases. For these, treatment is much more expensive, has lower efficacy, longer duration and higher toxicity [8]. Intensive research, as well as adapted guidelines to speed up the drug admission process, are necessary to keep up with resistance spread. We view both phage therapy and enzybiotics as potential contributors to future TB treatment, which would be applicable to both drug-susceptible and drug-resistant TB strains. Research should focus on exploiting phage effectors and derive more intelligent antibacterials to (i) target the pathogen only, not the microflora and (ii) exploit new mechanisms which circumvent evolved antibiotic resistance in MDR strains. These drugs could potentially be added to or even replace the current TB treatment regimen and contribute to WHO ‘End TB Strategy’.
- Summary: Studies of phage therapy against TB in animal models yielded promising results. Also, mycobacteriophage endolysin has been shown to effectively reduce TB infection [59]. However, the search for new phage effectors can be intensified, focussing on different mycobacterial molecular targets. The important groundwork was laid by projects initiated by Professor Hatfull: the sequencing of over 1720 mycobacteriophage genomes and their host spectrum characterisation [18] makes successive work to identify specific inhibitors much easier. Phages are intracellular pathogens to their bacterial hosts. In phage–host coevolution, continual development of resistance by mutation of host structures was continually overcome by new phage strategies. This race has been going on for thousands of years during which more and more powerful, adapted and specialised targeting was developed by the phage. Thus, it is not surprising that phages are highly specific to their host, and the molecular target can often not be exchanged by the homolog from another bacterial species. Selectivity and precision are clear advantages of phage gene products over antibiotics [44]. The study of molecular interactions between phage and host proteins during phage infection reveals that phages utilise various mechanism to target essential host processes, such as transcription. A number of studies show different mechanisms used by phages to hijack host transcription machinery for transcription of their own genes [76,81]. Remarkably, phage transcription inhibition differs from transcription inhibition by known drugs, thus it is likely that drug resistance does not affect the impact of the phage inhibitor. Thus cytotoxicity of phage gene products could potentially be exploited in MDR- and XDR-TB therapy.

- Future directions: Current research is focusing on DS- and DR-TB treatment from different angles, such as improvement of existing substances like rifampicin [83]. Research in other infectious bacterial diseases shows phage administration can stimulate the immune system and work in synergy with antibiotics [38–42], potentially attributing to phage lysins which destabilise the cell envelope to increase cytosolic drug concentration and activity of drugs which attack the cell envelope [52–54]. Lysins as enzybiotics have been shown to be bactericidal and support antibiotics. Also, mycobacteriophage endolysins are known [55,57]. Thus far, enzybiotics research is mainly focussed on phage endolysins. However, other phage gene products, which interact with essential cellular structures or processes, could be used as biocins as well.

Abbreviations

BCG, Bacille Calmette–Guerin; DR, drug-resistant; DR-TB, drug-resistant tuberculosis; DS-TB, drug-sensitive tuberculosis; Fdx, fidaxomicin; Lys, endolysin; MDR, multidrug resistance; MDR-TB, multidrug-resistant tuberculosis; MRSA, methicillin-resistant *Staphylococcus aureus*; Myx, myxopyronin; ORF, open reading frame; PAS, phage–antibiotic synergy; Rif, rifampicin; RNAP, RNA polymerase; RPc, RNAP-promoter closed complex; RPo, RNAP-promoter open complex; TB, tuberculosis; TEC, ternary elongation complex; XDR-TB, extensively drug-resistant tuberculosis; σ , sigma factor.

Competing Interests

The Authors declare that there are no competing interests associated with the manuscript.

References

- 1 WHO. (2018) *Global Tuberculosis Report 2018*, WHO, Geneva
- 2 Smith, I. (2003) Mycobacterium tuberculosis pathogenesis and molecular determinants of virulence. *Clin. Microbiol. Rev.* **16**, 463–496 <https://doi.org/10.1128/CMR.16.3.463-496.2003>
- 3 Hoffmann, C., Leis, A., Niederweis, M., Plitzko, J.M. and Engelhardt, H. (2008) Disclosure of the mycobacterial outer membrane: cryo-electron tomography and vitreous sections reveal the lipid bilayer structure. *Proc. Natl Acad. Sci. U.S.A.* **105**, 3963–3967 <https://doi.org/10.1073/pnas.0709530105>
- 4 Smith, T., Wolff, K.A. and Nguyen, L. (2013) Molecular biology of drug resistance in *Mycobacterium tuberculosis*. *Curr. Top. Microbiol. Immunol.* **374**, 53–80 https://doi.org/10.1007/82_2012_279
- 5 Hofman, S., Segers, M.M., Ghimire, S., Bolhuis, M.S., Sturkenboom, M.G.G., Van Soolingen, D. et al. (2016) Emerging drugs and alternative possibilities in the treatment of tuberculosis. *Expert Opin. Emerg. Drugs* **21**, 103–116 <https://doi.org/10.1517/14728214.2016.1151000>
- 6 Koch, A., Mizrahi, V. and Warner, D.F. (2014) The impact of drug resistance on *Mycobacterium tuberculosis* physiology: what can we learn from rifampicin? *Emerg. Microbes Infect.* **3**, e17 <https://doi.org/10.1038/emi.2014.17>
- 7 Zumla, A., Mwaba, P., Squire, S.B. and Grange, J.M. (1999) The tuberculosis pandemic—Which way now? *J. Infect.* **38**, 74–79 [https://doi.org/10.1016/S0163-4453\(99\)90072-5](https://doi.org/10.1016/S0163-4453(99)90072-5)
- 8 Tiberi, S., Muñoz-Torrico, M., Duarte, R., Dalcolmo, M., D'Ambrosio, L. and Migliori, G.B. (2018) New drugs and perspectives for new anti-tuberculosis regimens. *Pulmonology* **24**, 86–98 <https://doi.org/10.1016/j.rppnen.2017.10.009>
- 9 Udawadia, Z. and Vendoti, D. (2013) Totally drug-resistant tuberculosis (TDR-TB) in India: every dark cloud has a silver lining. *J. Epidemiol. Community Health* **67**, 471–472 <https://doi.org/10.1136/jech-2012-201640>
- 10 Stoffels, K., Allix-Béguec, C., Groenen, G., Wanlin, M., Berkvens, D., Mathys, V. et al. (2013) From multidrug- to extensively drug-resistant tuberculosis: upward trends as seen from a 15-year nationwide study. *PLoS ONE* **8**, e63128 <https://doi.org/10.1371/journal.pone.0063128>
- 11 Blumberg, H.M., Burman, W.J., Chaisson, R.E., Daley, C.L., Etkind, S.C., Friedman, L.N. et al. (2003) American thoracic society/centers for disease control and prevention/infectious diseases society of America. *Am. J. Respir. Crit. Care Med.* **167**, 603–662 <https://doi.org/10.1164/rccm.167.4.603>
- 12 Winston, C.A. and Mitruka, K. (2012) Treatment duration for patients with drug-resistant tuberculosis, United States. *Emerg. Infect. Dis.* **18**, 1201–1202 <https://doi.org/10.3201/eid1807.120261>
- 13 Vaudry, W. (2003) 'To BCG or not to BCG, that is the question!'. The challenge of BCG vaccination: why can't we get it right? *Paediatr. Child Health.* **8**, 141–144 <https://doi.org/10.1093/pch/8.3.141>
- 14 Dawson, R., Condos, R., Tse, D., Huie, M.L., Ress, S., Tseng, C.-H. et al. (2009) Immunomodulation with recombinant interferon- γ 1b in pulmonary tuberculosis. *PLoS ONE* **4**, e6984 <https://doi.org/10.1371/journal.pone.0006984>
- 15 Batdelger, D., Dandii, D., Jirathitikal, V. and Bourinbaier, A.S. (2008) Open-label trial of therapeutic immunization with oral V-5 Immunitor (V5) vaccine in patients with chronic hepatitis C. *Vaccine* **26**, 2733–2737 <https://doi.org/10.1016/j.vaccine.2008.03.021>
- 16 Arjanova O, V., Prihoda, N.D., Yurchenko L, V., Sokolenko, N.I., Frolov, V.M., Tarakanovskaya, M.G. et al. (2011) Adjunct oral immunotherapy in patients with re-treated, multidrug-resistant or HIV-coinfected TB. *Immunotherapy* **3**, 181–191 <https://doi.org/10.2217/imt.10.96>
- 17 Butov, D.A., Efremenko Y, V., Prihoda, N.D., Zaitzeva, S.I., Yurchenko L, V., Sokolenko, N.I. et al. (2013) Randomized, placebo-controlled phase II trial of heat-killed *Mycobacterium vaccae* (Immodulon batch) formulated as an oral pill (V7). *Immunotherapy* **5**, 1047–1054 <https://doi.org/10.2217/imt.13.110>

- 18 Phagesdb. Phagesdb.org. Available from: <https://phagesdb.org> [Accessed 22 March 2019]
- 19 Hatfull, G.F. (2014) Mycobacteriophages: windows into tuberculosis. *PLoS Pathog.* **10**, e1003953 <https://doi.org/10.1371/journal.ppat.1003953>
- 20 Jacobs-Sera, D., Marinelli, L.J., Bowman, C., Broussard, G.W., Guerrero Bustamante, C., Boyle, M.M. et al. (2012) On the nature of mycobacteriophage diversity and host preference. *Virology* **434**, 187–201 <https://doi.org/10.1016/j.virol.2012.09.026>
- 21 Young, R., Wang, I.-N. and Roof, W.D. (2000) Phages will out: strategies of host cell lysis. *Trends Microbiol.* **8**, 120–128 [https://doi.org/10.1016/S0966-842X\(00\)01705-4](https://doi.org/10.1016/S0966-842X(00)01705-4)
- 22 Nanda, A.M., Heyer, A., Krämer, C., Grünberger, A., Kohlheyer, D. and Frunzke, J. (2014) Analysis of SOS-induced spontaneous prophage induction in *Corynebacterium glutamicum* at the single-cell level. *J. Bacteriol.* **196**, 180–188 <https://doi.org/10.1128/JB.01018-13>
- 23 Echols, H. (1972) Developmental pathways for the temperate phage: lysis VS lysogeny. *Annu. Rev. Genet.* **6**, 157–190 <https://doi.org/10.1146/annurev.ge.06.120172.001105>
- 24 Fortier, L.-C. and Sekulovic, O. (2013) Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence* **4**, 354–365 <https://doi.org/10.4161/viru.24498>
- 25 Sula, L., Sulová, J. and Stolcpartová, M. (1981) Therapy of experimental tuberculosis in guinea pigs with mycobacterial phages DS-6A, GR-21T, My-327. *Czech. Med.* **4**, 209–214 PMID:7327068
- 26 Zemskova, Z.S. and Dorozhkova, I.R. (1991) Pathomorphological assessment of the therapeutic effect of mycobacteriophages in tuberculosis. *Probl. Tuberk.* **11**, 63–66 PMID:1775467
- 27 Trigo, G., Martins, T.G., Fraga, A.G., Longatto-Filho, A., Castro, A.G., Azeredo, J. et al. (2013) Phage therapy is effective against infection by *Mycobacterium ulcerans* in a murine footpad model. *PLoS Negl. Trop. Dis.* **7**, e2183 <https://doi.org/10.1371/journal.pntd.0002183>
- 28 Bessman, M.J., Walsh, J.D., Dunn, C.A., Swaminathan, J., Weldon, J.E. and Shen, J. (2001) The Gene ygdP, associated with the invasiveness of *Escherichia coli* K1, designates a nudix hydrolase, Orf176, active on adenosine (5′)-pentaphospho-(5′)-adenosine (Ap5A). *J. Biol. Chem.* **276**, 37834–8 <https://doi.org/10.1074/jbc.M107032200>
- 29 Chhibber, S., Kaur, S. and Kumari, S. (2008) Therapeutic potential of bacteriophage in treating *Klebsiella pneumoniae* B5055-mediated lobar pneumonia in mice. *J. Med. Microbiol.* **57**, 1508–1513 <https://doi.org/10.1099/jmm.0.2008/002873-0>
- 30 Capparelli, R., Parlato, M., Borriello, G., Salvatore, P. and Iannelli, D. (2007) Experimental phage therapy against *Staphylococcus aureus* in mice. *Antimicrob. Agents Chemother.* **51**, 2765–2773 <https://doi.org/10.1128/AAC.01513-06>
- 31 Röthlin, M. (1993) Intraoperative Cholangiographie und Sonographie bei der laparoskopischen Cholezystektomie. *Therapeutische. Umschau.* **50**, 553–558
- 32 Broxmeyer, L., Sosnowska, D., Miltner, E., Chacón, O., Wagner, D., McGarvey, J. et al. (2002) Killing of *Mycobacterium avium* and *Mycobacterium tuberculosis* by a mycobacteriophage delivered by a nonvirulent mycobacterium: a model for phage therapy of intracellular bacterial pathogens. *J. Infect. Dis.* **186**, 1155–1160 <https://doi.org/10.1086/343812>
- 33 Mankiewicz, E. and Béland, J.E. (1972) Failure to produce experimental sarcoidosis in Guinea pigs with *Mycobacterium tuberculosis* and mycobacteriophage DS6A. *Am. Rev. Respir. Dis.* **106**, 284–285 <https://doi.org/10.1164/arrd.1972.106.2.284>
- 34 Krut, O. and Bekeredjian-Ding, I. (2018) Contribution of the immune response to phage therapy. *J. Immunol.* **200**, 3037–3044 <https://doi.org/10.4049/jimmunol.1701745>
- 35 Singla, S., Harjai, K., Katare, O.P. and Chhibber, S. (2016) Encapsulation of bacteriophage in liposome accentuates its entry in to macrophage and shields it from neutralizing antibodies. *PLoS ONE* **11**, e0153777 <https://doi.org/10.1371/journal.pone.0153777>
- 36 Nieth, A., Verseux, C., Barnert, S., Süss, R. and Römer, W. (2015) A first step toward liposome-mediated intracellular bacteriophage therapy. *Expert Opin. Drug Deliv.* **12**, 1411–1424 <https://doi.org/10.1517/17425247.2015.1043125>
- 37 Leung, C.Y.J. and Weitz, J.S. (2017) Modeling the synergistic elimination of bacteria by phage and the innate immune system. *J. Theor. Biol.* **429**, 241–252 <https://doi.org/10.1016/j.jtbi.2017.06.037>
- 38 Roach, D.R., Leung, C.Y., Henry, M., Morello, E., Singh, D., Di Santo, J.P. et al. (2017) Synergy between the host immune system and bacteriophage is essential for successful phage therapy against an acute respiratory pathogen. *Cell Host Microbe* **22**, 38–47.e4 <https://doi.org/10.1016/j.chom.2017.06.018>
- 39 Kim, M., Jo, Y., Hwang, Y.J., Hong, H.W., Hong, S.S., Park, K. et al. (2018) Phage-antibiotic synergy via delayed lysis. *Appl. Environ. Microbiol.* **84**, e02085-18 <https://doi.org/10.1128/AEM.02085-18>
- 40 Jo, A., Ding, T. and Ahn, J. (2016) Synergistic antimicrobial activity of bacteriophages and antibiotics against *Staphylococcus aureus*. *Food Sci. Biotechnol.* **25**, 935–940 <https://doi.org/10.1007/s10068-016-0153-0>
- 41 Tkhalishvili, T., Lombardi, L., Klatt, A.-B., Trampuz, A. and Di Luca, M. (2018) Bacteriophage Sb-1 enhances antibiotic activity against biofilm, degrades exopolysaccharide matrix and targets persisters of *Staphylococcus aureus*. *Int. J. Antimicrob. Agents* **52**, 842–853 <https://doi.org/10.1016/j.ijantimicag.2018.09.006>
- 42 Ryan, E.M., Alkawareek, M.Y., Donnelly, R.F. and Gilmore, B.F. (2012) Synergistic phage-antibiotic combinations for the control of *Escherichia coli* biofilms *in vitro*. *FEMS Immunol. Med. Microbiol.* **65**, 395–398 <https://doi.org/10.1111/j.1574-695X.2012.00977.x>
- 43 Oliveira, A., Ribeiro, H.G., Silva, A.C., Silva, M.D., Sousa, J.C., Rodrigues, C.F. et al. (2017) Synergistic antimicrobial interaction between honey and phage against *Escherichia coli* biofilms. *Front. Microbiol.* **8**, 2407 <https://doi.org/10.3389/fmicb.2017.02407>
- 44 Hermoso, J.A., García, J.L. and García, P. (2007) Taking aim on bacterial pathogens: from phage therapy to enzybiotics. *Curr. Opin. Microbiol.* **10**, 461–472 <https://doi.org/10.1016/j.mib.2007.08.002>
- 45 São-José, C. (2018) Engineering of phage-derived lytic enzymes: improving their potential as antimicrobials. *Antibiotics* **7**, E29 <https://doi.org/10.3390/antibiotics7020029>
- 46 Domingo-Calap, P. and Delgado-Martínez, J. (2018) Bacteriophages: protagonists of a post-antibiotic era. *Antibiotics* **7**, E66 <https://doi.org/10.3390/antibiotics7030066>
- 47 Maxwell, K.L. (2017) Perspective the anti-CRISPR story : a battle for survival. *Mol. Cell* **68**, 8–14 <https://doi.org/10.1016/j.molcel.2017.09.002>
- 48 Hendrix, R.W. (2002) Bacteriophages: evolution of the majority. *Theor. Popul. Biol.* **61**, 471–480 <https://doi.org/10.1006/tpbi.2002.1590>
- 49 Rybníček, J., Plum, G., Robinson, N., Small, P.L. and Hartmann, P. (2008) Identification of three cytotoxic early proteins of mycobacteriophage L5 leading to growth inhibition in *Mycobacterium smegmatis*. *Microbiology* **154**, 2304–2314 <https://doi.org/10.1099/mic.0.2008/017004-0>

- 50 Hatfull, G.F. and Sarkis, G.J. (1993) DNA sequence, structure and gene expression of mycobacteriophage L5: a phage system for mycobacterial genetics. *Mol. Microbiol.* **7**, 395–405 <https://doi.org/10.1111/j.1365-2958.1993.tb01131.x>
- 51 Donnelly-Wu, M.K., Jacobs, W.R. and Hatfull, G.F. (1993) Superinfection immunity of mycobacteriophage L5: applications for genetic transformation of mycobacteria. *Mol. Microbiol.* **7**, 407–417 <https://doi.org/10.1111/j.1365-2958.1993.tb01132.x>
- 52 Schmelcher, M., Donovan, D.M. and Loessner, M.J. (2012) Bacteriophage endolysins as novel antimicrobials. *Fut. Microbiol.* **7**, 1147–1171 <https://doi.org/10.2217/fmb.12.97>
- 53 Nelson, D., Loomis, L. and Fischetti, V.A. (2001) Prevention and elimination of upper respiratory colonization of mice by group A streptococci by using a bacteriophage lytic enzyme. *Proc. Natl Acad. Sci. U.S.A.* **98**, 4107–4112 <https://doi.org/10.1073/pnas.061038398>
- 54 Bustamante, N., Campillo, N.E., García, E., Gallego, C., Pera, B., Diakun, G.P. et al. (2010) Cpl-7, a lysozyme encoded by a pneumococcal bacteriophage with a novel cell wall-binding motif. *J. Biol. Chem.* **285**, 33184–33196 <https://doi.org/10.1074/jbc.M110.154559>
- 55 Mahapatra, S., Piechota, C., Gil, F., Ma, Y., Huang, H., Scherman, M.S. et al. (2013) Mycobacteriophage Ms6 LysA: a peptidoglycan amidase and a useful analytical tool. *Appl. Environ. Microbiol.* **79**, 768–773 <https://doi.org/10.1128/AEM.02263-12>
- 56 Gil, F., Catalão, M.J., Moniz-Pereira, J., Leandro, P., McNeil, M. and Pimentel, M. (2008) The lytic cassette of mycobacteriophage Ms6 encodes an enzyme with lipolytic activity. *Microbiology* **154**, 1364–1371 <https://doi.org/10.1099/mic.0.2007/014621-0>
- 57 Catalão, M.J. and Pimentel, M. (2018) Mycobacteriophage lysis enzymes: targeting the mycobacterial cell envelope. *Viruses* **10**, E428 <https://doi.org/10.3390/v10080428>
- 58 Catalão, M.J., Milho, C., Gil, F., Moniz-Pereira, J. and Pimentel, M. (2011) A second endolysin gene is fully embedded in-frame with the lysA gene of mycobacteriophage Ms6. *PLoS ONE* **6**, e20515 <https://doi.org/10.1371/journal.pone.0020515>
- 59 Lai, M.-J., Liu, C.-C., Jiang, S.-J., Soo, P.-C., Tu, M.-H., Lee, J.-J. et al. (2015) Antimycobacterial activities of endolysins derived from a mycobacteriophage, BTCU-1. *Molecules* **20**, 19277–19290 <https://doi.org/10.3390/molecules201019277>
- 60 Mediavilla, J., Jain, S., Kriakov, J., Ford, M.E., Duda, R.L., Jacobs, W.R. et al. (2000) Genome organization and characterization of mycobacteriophage Bxb1. *Mol. Microbiol.* **38**, 955–970 <https://doi.org/10.1046/j.1365-2958.2000.02183.x>
- 61 Wang, Z., Kong, L., Liu, Y., Fu, Q., Cui, Z., Wang, J. et al. (2018) A phage lysin fused to a cell-penetrating peptide kills intracellular methicillin-resistant *Staphylococcus aureus* in keratinocytes and has potential as a treatment for skin infections in mice. *Appl. Environ. Microbiol.* **84**, e00380-18 <https://doi.org/10.1128/AEM.02085-18>
- 62 Cheng, Q., Nelson, D., Zhu, S. and Fischetti, V.A. (2005) Removal of group B streptococci colonizing the vagina and oropharynx of mice with a bacteriophage lytic enzyme. *Antimicrob. Agents Chemother.* **49**, 111–117 <https://doi.org/10.1128/AAC.49.1.111-117.2005>
- 63 Wall, R.J., Powell, A.M., Paape, M.J., Kerr, D.E., Bannerman, D.D., Pursel, V.G. et al. (2005) Genetically enhanced cows resist intramammary *Staphylococcus aureus* infection. *Nat. Biotechnol.* **23**, 445–451 <https://doi.org/10.1038/nbt1078>
- 64 Wei, L., Wu, J., Liu, H., Yang, H., Rong, M., Li, D. et al. (2013) A mycobacteriophage-derived trehalose-6,6'-dimycolate-binding peptide containing both antimycobacterial and anti-inflammatory abilities. *FASEB J.* **27**, 3067–3077 <https://doi.org/10.1096/fj.13-227454>
- 65 Teng, T., Liu, J. and Wei, H. (2015) Anti-mycobacterial peptides: from human to phage. *Cell Physiol. Biochem.* **35**, 452–466 <https://doi.org/10.1159/000369711>
- 66 Iona, E., Pardini, M., Mustazzolu, A., Piccaro, G., Nisini, R., Fattorini, L. et al. (2016) *Mycobacterium tuberculosis* gene expression at different stages of hypoxia-induced dormancy and upon resuscitation. *J. Microbiol.* **54**, 565–572 <https://doi.org/10.1007/s12275-016-6150-4>
- 67 Tabib-Salazar, A., Liu, B., Doughty, P., Lewis, R.A., Ghosh, S., Parsy, M.L. et al. (2013) The actinobacterial transcription factor RbpA binds to the principal sigma subunit of RNA polymerase. *Nucleic Acids Res.* **41**, 5679–5691 <https://doi.org/10.1093/nar/gkt277>
- 68 Sudalaiyadum Perumal, A., Vishwakarma, R.K., Hu, Y., Morichaud, Z. and Brodolin, K. (2018) RbpA relaxes promoter selectivity of *M. tuberculosis* RNA polymerase. *Nucleic Acids Res.* **46**, 10106–10118 <https://doi.org/10.1093/nar/gky714>
- 69 Imashimizu, M., Shimamoto, N., Oshima, T. and Kashlev, M. (2014) Transcription elongation. Heterogeneous tracking of RNA polymerase and its biological implications. *Transcription* **5**, e28285 <https://doi.org/10.4161/trns.28285>
- 70 Czyz, A., Mooney, R.A., Iaconi, A. and Landick, R. (2014) Mycobacterial RNA polymerase requires a U-tract at intrinsic terminators and is aided by NusG at suboptimal terminators. *mBio* **5**, e00931 <https://doi.org/10.1128/mBio.00931-14>
- 71 Botella, L., Vaujourgeix, J., Livny, J. and Schnappinger, D. (2017) Depleting *Mycobacterium tuberculosis* of the transcription termination factor Rho causes pervasive transcription and rapid death. *Nat. Commun.* **8**, 14731 <https://doi.org/10.1038/ncomms14731>
- 72 D'Heygère, F., Schwartz, A., Coste, F., Castaing, B. and Boudvillain, M. (2015) ATP-dependent motor activity of the transcription termination factor Rho from *Mycobacterium tuberculosis*. *Nucleic Acids Res.* **43**, 6099–6111 <https://doi.org/10.1093/nar/gkv505>
- 73 Nechaev, S. and Severinov, K. (2003) Bacteriophage-induced modifications of host RNA polymerase. *Annu. Rev. Microbiol.* **57**, 301–322 <https://doi.org/10.1146/annurev.micro.57.030502.090942>
- 74 Geiduschek, E.P. and Kassavetis, G.A. (2010) Transcription of the T4 late genes. *Viol. J.* **7**, 288 <https://doi.org/10.1186/1743-422X-7-288>
- 75 Ford, M.E., Stenstrom, C., Hendrix, R.W. and Hatfull, G.F. (1998) Mycobacteriophage TM4: genome structure and gene expression. *Tuber. Lung Dis.* **79**, 63–73 <https://doi.org/10.1054/tuld.1998.0007>
- 76 Rybníček, J., Nowag, A., Van Gumpel, E., Nissen, N., Robinson, N., Plum, G. et al. (2010) Insights into the function of the WhiB-like protein of mycobacteriophage TM4 - A transcriptional inhibitor of WhiB2. *Mol. Microbiol.* **77**, 642–657 <https://doi.org/10.1111/j.1365-2958.2010.07235.x>
- 77 Fan, X., Duan, X., Tong, Y., Huang, Q., Zhou, M., Wang, H. et al. (2016) The global reciprocal reprogramming between mycobacteriophage SWU1 and mycobacterium reveals the molecular strategy of subversion and promotion of phage infection. *Front. Microbiol.* **7**, 41 <https://doi.org/10.3389/fmicb.2016.00041>
- 78 Oldfield, L.M. and Hatfull, G.F. (2014) Mutational analysis of the mycobacteriophage BPs promoter PR reveals context-dependent sequences for mycobacterial gene expression. *J. Bacteriol.* **196**, 3589–3597 <https://doi.org/10.1128/JB.01801-14>
- 79 Santangelo, T.J. and Artsimovitch, I. (2011) Termination and antitermination: RNA polymerase runs a stop sign. *Nat. Rev. Microbiol.* **9**, 319–329 <https://doi.org/10.1038/nrmicro2560>
- 80 Gottesman, M.E. and Weisberg, R.A. (1995) Termination and antitermination of transcription in temperate bacteriophages. *Semin. Virol.* **6**, 35–42 [https://doi.org/10.1016/S1044-5773\(05\)80007-1](https://doi.org/10.1016/S1044-5773(05)80007-1)

- 81 Garcia, M., Pimentel, M. and Moniz-Pereira, J. (2002) Expression of mycobacteriophage Ms6 lysis genes is driven by two sigma(70)-like promoters and is dependent on a transcription termination signal present in the leader RNA. *J. Bacteriol.* **184**, 3034–3043 <https://doi.org/10.1128/JB.184.11.3034-3043.2002>
- 82 Ghosh, G., Reddy, J., Sambhare, S. and Sen, R. (2018) A bacteriophage capsid protein is an inhibitor of a conserved transcription terminator of various bacterial pathogens. *J. Bacteriol.* **200**, 1–16 <https://doi.org/10.1128/JB.00380-17>
- 83 Mosaei, H., Molodtsov, V., Kepplinger, B., Harbottle, J., Moon, C.W., Jeeves, R.E. et al. (2018) Mode of action of Kanglemycin A, an Ansamycin natural product that is active against Rifampicin-resistant *Mycobacterium tuberculosis*. *Mol. Cell* **72**, 263–274.e5 <https://doi.org/10.1016/j.molcel.2018.08.028>
- 84 Orsini, G., Ouhammouch, M., Le Caer, J.P. and Brody, E.N. (1993) The AsiA gene of bacteriophage T4 codes for the anti-sigma 70 protein. *J. Bacteriol.* **175**, 85–93 <https://doi.org/10.1128/jb.175.1.85-93.1993>
- 85 Pal, D., Vuthoori, M., Pande, S., Wheeler, D. and Hinton, D.M. (2003) Analysis of regions within the bacteriophage T4 AsiA protein involved in its binding to the σ 70 subunit of *E. coli* RNA polymerase and its role as a transcriptional inhibitor and co-activator. *J. Mol. Biol.* **325**, 827–841 [https://doi.org/10.1016/S0022-2836\(02\)01307-4](https://doi.org/10.1016/S0022-2836(02)01307-4)
- 86 Lambert, L.J., Wei, Y., Schirf, V., Demeler, B. and Werner, M.H. (2004) T4 AsiA blocks DNA recognition by remodeling sigma70 region 4. *EMBO J.* **23**, 2952–2962 <https://doi.org/10.1038/sj.emboj.7600312>
- 87 Ma, C., Yang, X. and Lewis, P.J. (2016) Bacterial transcription inhibitor of RNA polymerase holoenzyme formation by structure-based drug design: from in silico screening to validation. *ACS Infect. Dis.* **2**, 39–46 <https://doi.org/10.1021/acsinfecdis.5b00058>
- 88 Dehbi, M., Moeck, G., Arhin, F.F., Bauda, P., Bergeron, D., Kwan, T. et al. (2009) Inhibition of transcription in *Staphylococcus aureus* by a primary sigma factor-binding polypeptide from phage G1. *J. Bacteriol.* **191**, 3763–3771 <https://doi.org/10.1128/JB.00241-09>
- 89 Tagami, S., Sekine, S., Minakhin, L., Esyunina, D., Akasaka, R., Shirouzu, M. et al. (2014) Structural basis for promoter specificity switching of RNA polymerase by a phage factor. *Genes Dev.* **28**, 521–531 <https://doi.org/10.1101/gad.233916.113>
- 90 Ooi, W.-Y., Murayama, Y., Mekler, V., Minakhin, L., Severinov, K., Yokoyama, S. et al. (2018) A Thermus phage protein inhibits host RNA polymerase by preventing template DNA strand loading during open promoter complex formation. *Nucleic Acids Res.* **46**, 431–441 <https://doi.org/10.1093/nar/gkx1162>
- 91 Ouhammouch, M., Adelman, K., Harvey, S.R., Orsini, G. and Brody, E.N. (1995) Bacteriophage T4 MotA and Asia proteins suffice to direct *Escherichia coli* RNA polymerase to initiate transcription at T4 middle promoters. *Proc. Natl Acad. Sci. U.S.A.* **92**, 1451–1455 <https://doi.org/10.1073/pnas.92.5.1451>
- 92 Hinton, D.M., March-Amegadzie, R., Gerber, J.S. and Sharma, M. (1996) Bacteriophage T4 middle transcription system: T4-modified RNA polymerase; AsiA, a σ 70 binding protein; and transcriptional activator MotA. *Methods Enzymol.* **274**, 43–57 [https://doi.org/10.1016/S0076-6879\(96\)74007-7](https://doi.org/10.1016/S0076-6879(96)74007-7)
- 93 Guild, N., Gayle, M., Sweeney, R., Hollingsworth, T., Modeer, T. and Gold, L. (1988) Transcriptional activation of bacteriophage T4 middle promoters by the motA protein. *J. Mol. Biol.* **199**, 241–258 [https://doi.org/10.1016/0022-2836\(88\)90311-7](https://doi.org/10.1016/0022-2836(88)90311-7)
- 94 Osmundson, J. and Darst, S.A. (2013) Biochemical insights into the function of phage G1 gp67 in *Staphylococcus aureus*. *Bacteriophage* **3**, e24767 <https://doi.org/10.4161/bact.24767>
- 95 Nechaev, S. and Severinov, K. (1999) Inhibition of *Escherichia coli* RNA polymerase by bacteriophage T7 gene 2 protein. *J. Mol. Biol.* **289**, 815–826 <https://doi.org/10.1006/jmbi.1999.2782>
- 96 Klimuk, E., Akulenko, N., Makarova, K.S., Ceysens, P.-J., Volchenkov, I., Lavigne, R. et al. (2013) Host RNA polymerase inhibitors encoded by ϕ KMV-like phages of *Pseudomonas*. *Virology* **436**, 67–74 <https://doi.org/10.1016/j.virol.2012.10.021>
- 97 Fruth, M., Plaza, A., Hinsberger, S., Sahner, J.H., Hauptenthal, J., Bischoff, M. et al. (2014) Binding mode characterization of novel RNA polymerase inhibitors using a combined biochemical and NMR approach. *ACS Chem. Biol.* **9**, 2656–2663 <https://doi.org/10.1021/cb5005433>
- 98 Berdygulova, Z., Westblade, L.F., Florens, L., Koonin E, V., Chait, B.T., Ramanculov, E. et al. (2011) Temporal regulation of gene expression of the *Thermus thermophilus* bacteriophage P23-45. *J. Mol. Biol.* **405**, 125–142 <https://doi.org/10.1016/j.jmb.2010.10.049>
- 99 McClure, W.R. and Cech, C.L. (1978) On the mechanism of rifampicin inhibition of RNA synthesis. *J. Biol. Chem.* **253**, 8949–8956 PMID:363713
- 100 Campbell, E.A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A. et al. (2001) Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell* **104**, 901–912 [https://doi.org/10.1016/S0092-8674\(01\)00286-0](https://doi.org/10.1016/S0092-8674(01)00286-0)
- 101 Zhang, Y., Degen, D., Ho, M.X., Sineva, E., Ebright, K.Y., Ebright, Y.W. et al. (2014) GE23077 binds to the RNA polymerase 'i' and 'i+1' sites and prevents the binding of initiating nucleotides. *eLife* **3**, e02450 <https://doi.org/10.7554/eLife.02450>
- 102 Degen, D., Feng, Y., Zhang, Y., Ebright, K.Y., Ebright, Y.W., Gigliotti, M. et al. (2014) Transcription inhibition by the depsipeptide antibiotic salinamide A. *eLife* **3**, e02451 <https://doi.org/10.7554/eLife.02451>
- 103 Temiakov, D., Zenkin, N., Vassilyeva, M.N., Perederina, A., Tahirov, T.H., Kashkina, E. et al. (2005) Structural basis of transcription inhibition by antibiotic streptolydigin. *Mol. Cell* **19**, 655–666 <https://doi.org/10.1016/j.molcel.2005.07.020>
- 104 Zenkin, N., Severinov, K. and Yuzenkova, Y. (2015) Bacteriophage Xp10 anti-termination factor p7 induces forward translocation by host RNA polymerase. *Nucleic Acids Res.* **43**, 6299–6308 <https://doi.org/10.1093/nar/gkv586>
- 105 Adelman, K., Yuzenkova, J., La Porta, A., Zenkin, N., Lee, J., Lis, J.T. et al. (2004) Molecular mechanism of transcription inhibition by peptide antibiotic microcin J25. *Mol. Cell* **14**, 753–762 <https://doi.org/10.1016/j.molcel.2004.05.017>
- 106 Roberts, J.W. (1969) Termination factor for RNA synthesis. *Nature* **224**, 1168–1174 <https://doi.org/10.1038/2241168a0>
- 107 Yang, H., Ma, Y., Wang, Y., Yang, H., Shen, W. and Chen, X. (2014) Transcription regulation mechanisms of bacteriophages: recent advances and future prospects. *Bioengineered* **5**, 300–304 <https://doi.org/10.4161/bioe.32110>
- 108 Yuzenkova, J., Nechaev, S., Berlin, J., Rogulja, D., Kuznedelov, K., Inman, R. et al. (2003) Genome of *Xanthomonas oryzae* bacteriophage Xp10: an odd T-odd phage. *J. Mol. Biol.* **330**, 735–748 [https://doi.org/10.1016/S0022-2836\(03\)00634-X](https://doi.org/10.1016/S0022-2836(03)00634-X)
- 109 Berdygulova, Z., Esyunina, D., Miropolskaya, N., Mukhamedyarov, D., Kuznedelov, K., Nickels, B.E. et al. (2012) A novel phage-encoded transcription antiterminator acts by suppressing bacterial RNA polymerase pausing. *Nucleic Acids Res.* **40**, 4052–4063 <https://doi.org/10.1093/nar/gkr1285>

Metabolic cofactors act as initiating substrates for primase and affect replication primer processing

Christina Julius and Yulia Yuzenkova*

Centre for Bacterial Cell Biology, Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne, NE2 4AX, UK

*Corresponding author:

Yulia Yuzenkova, PhD

Centre for Bacterial Cell Biology

Institute for Cell and Molecular Biosciences,

Newcastle University,

Newcastle upon Tyne, NE2 4AX, UK

Phone: +44(0)1912083227

FAX: +44(0)1912083205

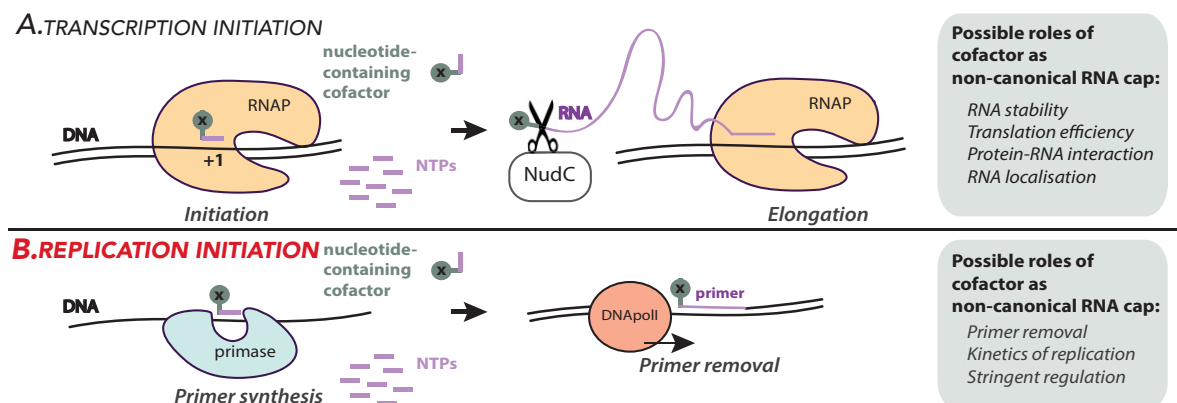
E-mail: y.yuzenkova@ncl.ac.uk

Abstract

Recently a new, non-canonical type of 5'-RNA capping with cellular metabolic cofactors was discovered in bacteria and eukaryotes. This type of capping is performed by RNA polymerases, the main enzymes of transcription, which initiate RNA synthesis with cofactors. Here we show that primase, the enzyme of replication which primes synthesis of DNA by making short RNA primers, initiates synthesis of replication primers using the number of metabolic cofactors. Primase DnaG of *E. coli* starts synthesis of RNA with cofactors NAD^+/NADH , FAD and DP-CoA *in vitro*. This activity does not affect primase specificity of initiation. ppGpp, the global starvation response regulator, strongly inhibits the non-canonical initiation by DnaG. Amino acid residues of a “basic ridge” define the binding determinant of cofactors to DnaG. Likewise, the human primase catalytic subunit P49 can use modified substrate m^7GTP for synthesis initiation.

For correct genome duplication, the RNA primer needs to be removed and Okazaki fragments ligated. We show that the efficiency of primer processing by DNA polymerase I is strongly affected by cofactors on the 5'-end of RNA. Overall our results suggest that cofactors at the 5' position of the primer influence regulation of initiation and Okazaki fragments processing.

Visual abstract



A. Non-canonical capping of RNA by RNA polymerase. RNA polymerase uses cellular cofactor as initiating substrate for RNA synthesis, instead of NTP. Then RNA chain grows, while cofactor remains attached and serves as cap. **B. Proposed mechanism of non-canonical initiation of RNA primer synthesis by DnaG primase during replication.** DnaG primase initiates synthesis of the primer for DNA replication using cellular cofactor. Primer stays annealed with the DNA template. DNApoll encounters cofactor, which affects the removal of primer.

Introduction

This work was inspired by the recent discovery of the non-canonical RNA capping phenomenon. Many RNA species in bacteria and eukaryotes bear metabolic adenine-containing cofactors at their 5'-end; NAD⁺ (nicotinamide adenine dinucleotide) and DP-CoA (diphospho- coenzyme A) and FAD (flavin adenine dinucleotide), as well as cell wall precursors UDP-GlcNAc (uridine 5'-diphospho-N-acetylglucosamine) and UDP-Glc (uridine 5'-diphosphate glucose) (1, 2). Unlike the classic cap m⁷G, non-canonical caps are installed by the main enzyme of transcription, RNA polymerase (RNAP)(3). It happens during initiation of transcription in a template-dependent manner – ADP-containing cofactors are incorporated at promoters with +1A start sites, i.e. promoters dictating ATP as the initiating substrate (3, 4), UDP-containing cell wall precursors - on +1U promoters (3).

By analogy with the classic cap, there are decapping enzymes for non-canonical caps, in *E. coli* it is NudC (NADH pyrophosphohydrolase of the NUDIX family). NudC processes NADylated RNAs into a monophosphorylated species that are quickly degraded in the cell (5).

Overall it appears that unrelated multi-subunit eukaryotic and bacterial RNAPs as well as the single-subunit RNAPs of mitochondria and viruses can utilize non-canonical initiating substrates (NCISs) and perform RNA capping (6). Another DNA-dependent enzyme initiating *de novo* synthesis of RNA is primase (DnaG in bacteria) which makes primers for replication and present in all organisms. It is not structurally related to either single subunit e.g. mitochondrial RNAP or multi subunit bacteria or eukaryotic RNAPs. In *E. coli*, DnaG recognises a consensus GTC motif and makes a 10-12 nucleotides long RNA primer. Primase acts in concert with other replication proteins. Primase requires DnaB helicase to start synthesis on double-stranded DNA (7). Primase is then displaced and the primer is elongated by the DNA polymerase III to produce the RNA/DNA polynucleotide of a leading strand or an Okazaki fragment of a lagging strand (8). RNA primers need to be removed via the combined actions of DNA polymerase I (PolI) and/or RNaseH before the Okazaki fragments and leading strand are ligated to complete genome replication. Thus, synthesis of an RNA primer by primase is believed to be a rate limiting step of replication (9), tightly coupled to other steps of the replication process. Primase plays a key role during assembly of the replisome (10), regulation of replication

elongation and Okazaki fragments length (11, 12) in both bacterial and eukaryotic systems. In bacteria, DnaG primase plays a part in a global concerted response to starvation via starvation alarmone ppGpp which inhibits primer synthesis in nutrient limited conditions (13).

Here we show that *E. coli* primase DnaG starts synthesis of a replication primer using a number of ADP-containing cofactors *in vitro*, including NAD⁺/ NADH, FAD and DP-coA. This reaction requires amino acid residues of the DnaG “basic ridge” region, and is inhibited by global starvation alarmone ppGpp. We also show that cofactors on the 5'-end of RNA specifically and differentially affect processing of this RNA by DNA polI. Our data suggest that 5'-cofactors influence initiation efficiency and the rate of processing of replication intermediates.

Results

***E. coli* DnaG primase initiates RNA primer synthesis using NAD⁺, NADH, FAD and DP-coA.**

DnaG primase functions as a low-processive RNA polymerase able to start *de novo* RNA synthesis on a DNA. We wanted to test if primase can initiate synthesis using ADP-containing metabolic cofactors, (structures on Fig. 1A), by analogy with other polymerases.

We used a general priming system, i.e. minimal system in the absence of single-strand DNA-binding protein. This set-up requires only DnaG for RNA primer synthesis on the short single-stranded DNA template containing GTC recognition motif (scheme on Fig. 1B)(7).

We found that DnaG makes a 13nt long RNA product in a subset of NTPs using either ATP, NAD⁺, NADH, FAD and DP-coA (Fig. 1B). To avoid confusion and for simplicity, we will refer to the length of RNA with conventional substrates even though cofactors are dinucleotides. Notably, DnaG incorporates NAD⁺ much less efficiently than NADH, in contrast to other RNAPs, which do not discriminate between NAD⁺ and NADH (4). We found that affinity to the non-canonical initiation substrates is comparable with their physiological concentrations. Michaelis constants we measured for ATP, NADH and FAD as initiating substrates, were 46.6, 109 and 390 μ M,

correspondingly. In actively growing in rich media *E. coli* cells, concentrations of ATP, NADH and FAD are 9.6 mM, 100-1000 μ M and 200 μ M, correspondingly (14, 15).

NADH capped primers were not susceptible to decapping by NudC, unless primase was washed away with high salt buffer, like in the experiment shown on Figure 1C. This result is in agreement with the view that full length primer stays bound to the DNA template and in complex with primase (12), in our case even if the primer contains extra moiety on its 5'-end.

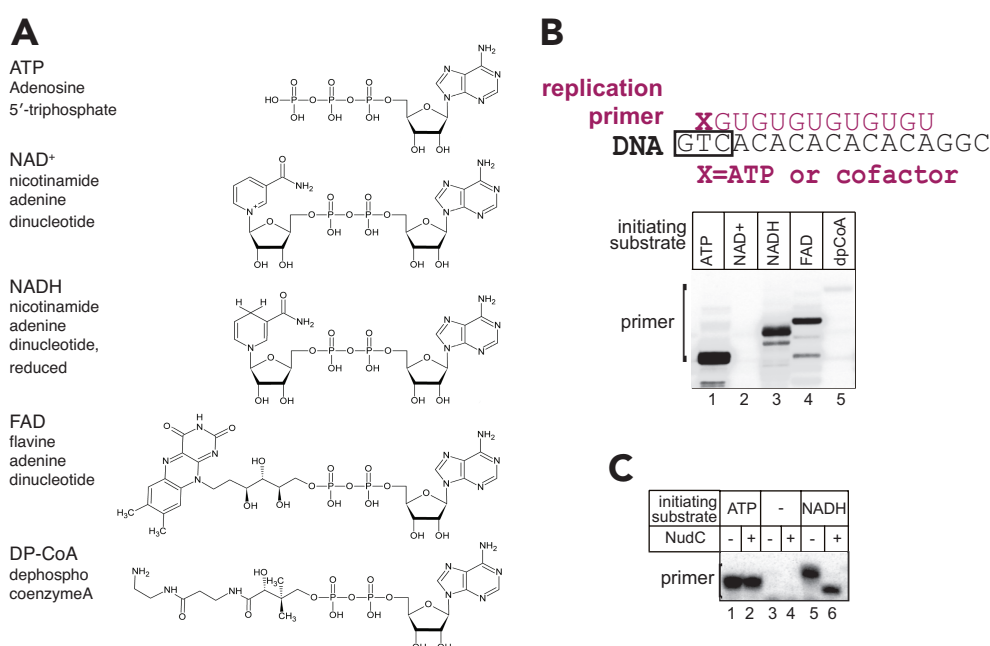
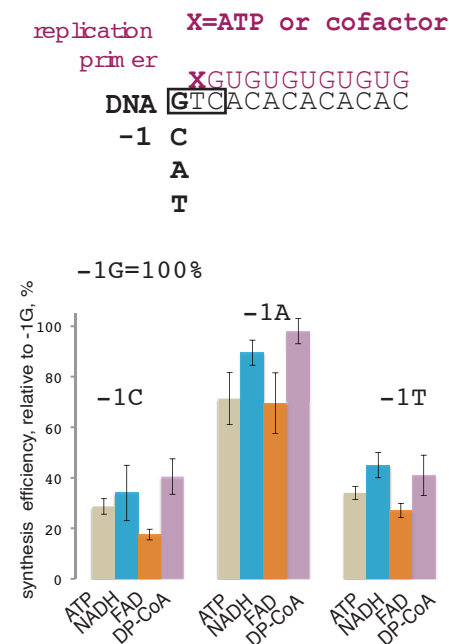


Figure 1. Primase DnaG initiates replication primer synthesis with NAD⁺, NADH, FAD and DP-coA. **A.** Structures of cofactors molecules (NAD⁺, NADH, FAD and DP-coA) in comparison to ATP, the preferred initiating substrate of primase. **B.** Replication primer synthesis on single-stranded DNA template, scheme above, with ATP, NAD⁺, NADH, FAD, and DP-coA as initiation substrates. **C.** RNA primer with 5'-NADH is susceptible for cleavage by NudC nuclease after DnaG was removed from the complex with high salt wash. Note that absence of ATP from the reaction results in an increased amount of non-specific product, which is not susceptible to NudC in lanes 5, 6. We assume that this band results from initiation with GTP present in the reaction.

Interestingly, after primase was removed, 5'-NADH of “naked” RNA primer annealed to DNA was susceptible to NudC, in contrast to the published result with 5'-NADH of RNA annealed to RNA (5). Therefore, NudC may have a limited window of opportunity to remove the cap from the primer during active replication and participate in processing of replication intermediates.

Cofactors substrates do not affect DnaG specificity of initiation

Cofactors are dinucleotides, and therefore can make additional contacts with DNA template at -1 position relative to the start site (scheme on Fig. 2). These contacts may affect initiation specificity by DnaG which recognises GTC motif on a template strand. The central T codes for A and G is a template -1 base. To test if cofactors incorporation depends on the nature of -1 base, we tested synthesis of RNA13 on templates with -1



changed to three other bases. This experiment we performed with initiation substrates concentrations roughly in the region of their corresponding K_m s (50 μ M for ATP and 500 μ M for cofactors). We found that primase preferred purines in this position, and the least preferred base is C (Fig. 2). Initiation with cofactors did not change these preferences, suggesting that cofactors do not make specific contacts with -1 base of the template. This result also implies that cofactors as substrates do not change specificity of DnaG initiation and do not lead to spurious initiation.

Figure 2. Cofactors do not make specific contacts with -1 DNA template base.

Synthesis efficiency of RNA13 started with ATP or NADH, FAD or DP-CoA on DNA templates with C, A or T at position -1 was compared to consensus -1G template. Relative efficiency of synthesis is shown in percentage from -1G template, error bars reflect standard deviations from three independent experiments.

Initiation of RNA synthesis with cofactors requires basic ridge amino acid residues of DnaG

Since cofactors do not make strong contacts with DNA, they most probably contact DnaG protein itself. A number of amino acid residues, including several in a “basic ridge”, were implicated in initiation nucleotide binding based on sequence conservation amongst primases and structural information for the *Staphylococcus aureus* primase (13). We tested synthesis of a primer by DnaG with amino acid substitutions, K229A, Y230A, K241A (basic ridge) and D309A (participating in metal chelation), which were all previously shown to influence initiating substrate incorporation (13, 16), in the presence of DnaB. We found that “basic ridge” substitutions K229A, Y230A, and to some extent K241A specifically inhibited capping with NADH and FAD (Fig. 3), in contrast with

D309A. We assumed that NMN and FMN moieties of the corresponding cofactors might make contacts with these amino acid residues either during binding of the initiating substrates or during the very first step of RNA synthesis.

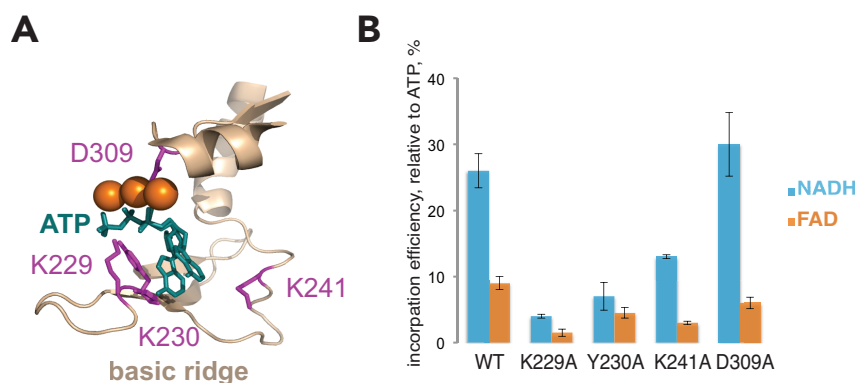
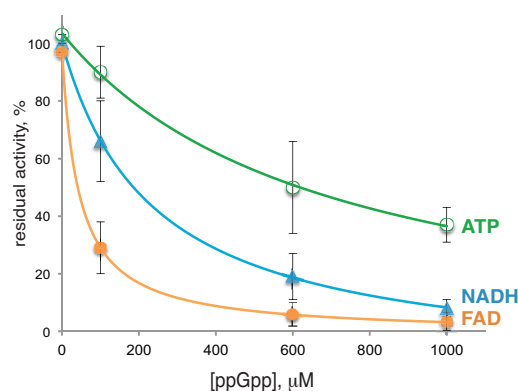


Figure 3. DnaG basic ridge residues affect initiation with NADH and FAD. A. Structure of the *S. aureus* DnaG-ATP complex in the vicinity of a catalytic site PDB 4EDG. ATP (in two conformations) is in teal, amino acid residues, which we tested for efficiency of cofactor initiation are in magenta. **B.** Relative efficiency of NADH and FAD utilisation as initiation substrates in comparison to that of ATP, for WT DnaG and mutants with amino acid residues substitutions to alanines at positions 229, 230, 241 and 309. Error bars reflect standard deviations from three independent experiments.

ppGpp strongly inhibits non-canonical initiation by DnaG.

Under nutrient deficient conditions, replication is inhibited via the action of global stringent response regulator, alarmone ppGpp, on DnaG. PpGpp binds DnaG at the active site (13), presumably overlapping with the binding site where caps bind during the initiation of primer synthesis. To test if ppGpp competes with capping cofactors NADH and FAD during initiation we measured the maximal rate of RNA product formation in the presence of increasing concentrations of ppGpp (control reactions were done with ATP as initiator). From the plot of residual activities vs ppGpp concentration, it appears that ppGpp competes with cofactors more efficiently than with ATP (Fig. 4). This tendency correlates with Michaelis constants for the initiation substrates which increase in a sequence ATP-NADH-FAD. Therefore, in the absence of structural information on DnaG complexes with cofactors, we can only suggest that binding site is shared by ppGpp and cofactors. During the stringent response, ppGpp concentration might raise above 150 μ M in *E. coli* (17) which is roughly, judging from plot, above or close to ppGpp inhibition constants for synthesis reaction of FAD-RNA and NADH-RNA. This

result might additionally mean that *in vivo* replication initiation with cofactors would be



inhibited earlier during transition into stationary phase or nutritional downshift.

Figure 4. Initiation with cofactors is more susceptible to inhibition by ppGpp than initiation with ATP. Residual DnaG activities are plotted against an increasing concentrations of ppGpp, in percentages from amounts in the absence of ppGpp. Error bars reflect standard deviations obtained from three independent experiments.

Cofactor 5'-caps differentially affect primer processing by PolI.

To complete replication, the leading strand and Okazaki fragments of a lagging strand need to be processed and ligated. The processing involves RNA primer degradation coupled to extension of previous Okazaki fragment by PolI, which possesses 5'-exonuclease and 3'- DNA polymerase activities. We examined if the presence of cofactor on the 5'-end of a RNA primer affects its removal by PolI. This experiment was done using a DNA-RNA construct mimicking part of replication intermediate (Fig. 5, top). The construct consisted of hairpin-containing DNA template with RNA primer (RNA12) produced by DnaG annealed to the single stranded DNA part. DnaG primase was subsequently removed by washing with high salt containing buffer. The DNA-RNA construct was immobilised on streptavidin beads via biotin on DNA, which ensured that the processing observed happens on an RNA annealed to a DNA (See Methods section for more details). Addition of PolI and dNTPs led to a stepwise degradation of RNA, as seen from kinetics of RNA12 degradation on a gel image and the plot below it on Fig. 5. PolI activity was stimulated by the 5'- end NADH; from Fig. 5 it can be seen that full length NADH-RNA12 disappeared in 30 seconds. In contrast, presence of 5'-FAD on the RNA12 greatly reduced the rate of degradation, and FAD-RNA12 was still visible after 4 minutes. ATP-RNA12 was degraded with somewhat intermediate rate. Our data suggest that 5' cofactors on the primer may influence its removal *in vivo*.

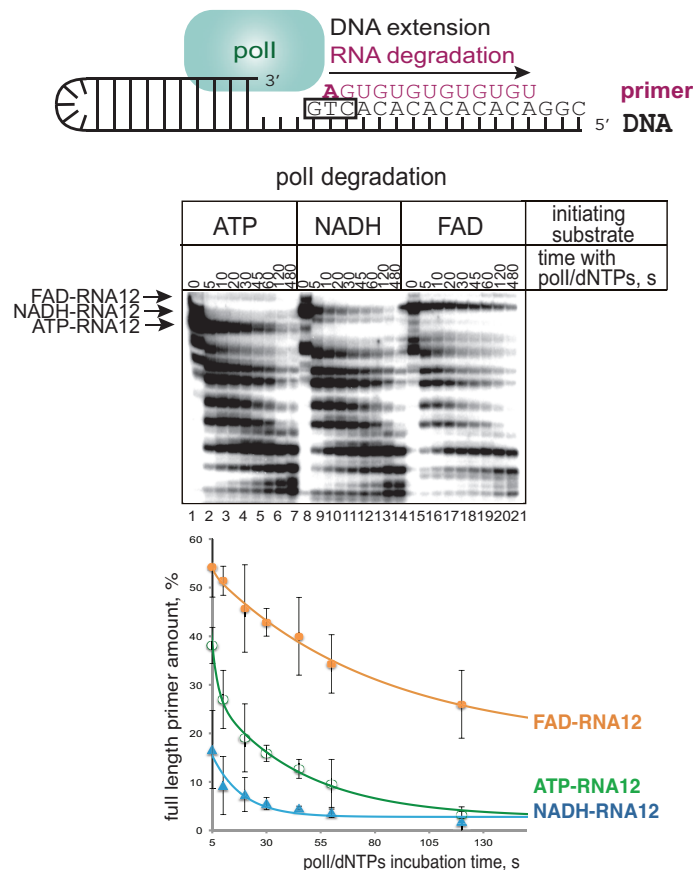


Figure 5. Cofactors at the 5'- end of a primer differentially affect its processing by PolI - NADH speeds it up, FAD inhibits. Scheme of the hairpin DNA substrate with annealed RNA made by DnaG is shown above the gel. On the gel products of time dependent degradation of the ATP-RNA12, NADH-RNA12 FAD-RNA12 made with either ATP, NADH or FAD as initiating substrates is shown. Below the gel image, the same degradation kinetics are presented on a plot. Plot shows amount of the initial full-length of primer as a function of incubation time with PolI and dNTPs. Error bars reflect standard deviations obtained from three independent experiments.

Eukaryotic primase catalytic subunit P49 uses modified initiating substrate

In addition to prokaryotic system, we wanted to test if the human primase catalytic subunit P49 utilises ADP containing cofactors. We analysed the formation of the first dinucleotide product by this enzyme on the single stranded DNA template (sequence on Fig. 6). We were unable to make this enzyme to start synthesis with ATP at a specific start site, and initiation with GTP was very inefficient on this template (lane 2). Yet, P49 efficiently produced a dinucleotide using m⁷GTP as initiating and UTP as the substrate for the second position (lane 4). This ready incorporation of a modified substrate hints at P49 general low fidelity and propensity to non-canonical initiation. Therefore, our result suggests that P49 might utilise a variety of non-canonical substrates with possible consequences for initiation kinetics and elongation to full length primer (18).

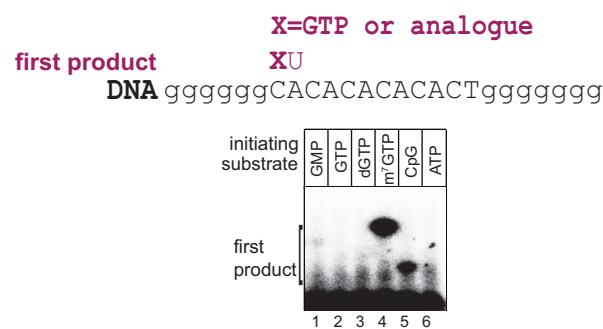


Figure 6. Human primase catalytic subunit P49 efficiently utilises m⁷GTP as initiating substrate. Scheme of the single-stranded DNA template is above the gel image. On the gel products of first step of synthesis with different initiating substrates and UTP as extending nucleotide are loaded.

Discussion

Here we showed that DnaG primase initiates synthesis of a replication primer with ADP-containing cellular cofactors. This ability is reminiscent of that of RNA polymerases of transcription, yet there are notable differences. We found that DnaG uses NAD⁺ very inefficiently compared to NADH, in striking contrast to bacterial and mitochondrial transcriptases which do not distinguish between the two (4, 6). This feature of DnaG might connect priming of replication to the redox state of the cell. We also showed that the template base at -1 position does not play role in non-canonical substrate utilisation by DnaG, unlike transcriptases, which at least in some instances are sensitive to the -1 base (3). This result suggests that the incorporation of cofactors does not lead to spurious initiation or increased noise in the system.

For the first time here we show that global starvation response regulator, ppGpp affects non-canonical initiation of RNA synthesis. It remains to be seen if incorporation of non-canonical substrates by bacterial RNAP is influenced by ppGpp.

RNA pieces of Okazaki fragments are destined to be removed. Yet despite the transient nature of these RNAs, a balance between kinetics of removal and extension defines the mean size and length distribution of Okazaki fragments and ultimately replication kinetics (9). We showed that the rate of processing of replication primer by PolII is affected specifically by 5'-cofactors. NADH greatly stimulates, and FAD and DP-CoA inhibit the processing. We also found that decapping nuclease NudC can remove 5'-cofactors from RNA even if RNA is annealed to DNA. Therefore, hypothetically, NudC may assist removal of primers aberrantly left unprocessed in the cell.

The propensity of the human primase catalytic subunit to incorporate efficiently a modified substrate, an analogue of the classic cap m⁷GTP, suggests the probability of non-canonical initiating of replication in eukaryotes.

The evidence on physiological roles and the regulatory mechanisms of cofactors as non-canonical initiating substrates in transcription are emerging. It was demonstrated that capping influences the stability of RNA (2), and capping stimulates RNAP escape from the promoter (3). We show that replication may also be affected by cofactor incorporation. Currently, more potential nucleotide analogues, including dinucleotide polyphosphates incorporated into 5' position of cellular RNAs are being discovered in both kingdoms (2). The role of these emerging substrates potentially extends to replication regulation.

Our results strongly suggest that cofactor initiation of replication happens *in vivo*, and future research would confirm this. At present we were unable to detect presence of cofactors on RNA primers *in vivo*, most probably due to the transient nature of a replication primer, relatively low number of primers per cell, and the sensitivity limitation of the methods we used. Nevertheless, we keep trying.

Materials and Methods

Strains and Plasmids

DnaG wt was overexpressed from pCA24N-dnaG plasmid (KAYO collection), DnaB wt from pCA24N-dnaB (strain JW4012) ASKA collection. DnaG was transferred to plasmid pET28a via Gibson Assembly. DnaG mutants K229A, Y230A, F238A, K241A and D309A were generated by Quick change mutagenesis from pET28a-dnaG wt basis.

Media and selection

E. coli were always grown in Luria bertani (LB) medium (broth or agar). In presence of pET28a Kanamycin (50 mg/L) was given to select for plasmid, pCA24N was selected with Chloramphenicol (34 mg/L).

Cloning of WT and mutant DnaG

Gibson Assembly and QuickChange protocols were used according to manufacturer's specifications and transformed into *E. coli* DH5 α cells. After sequencing, the plasmid was retransformed into *E. coli* T7 overexpression strain.

Protein purification

DnaG wt and mutants were grown in 1 L LB with Kanamycin to an OD₆₀₀= 0.5 at 37° C before induction with IPTG (1 mM) and continued growth at room temperature overnight. Cells were pelleted and stored at -80° C until lysis using sonication in Grinding Buffer (20 mM Tris-HCl pH 7.9, 200 mM NaCl, 5% Glycerol) and Ni-column

purification (His-trap) in Purification Buffers A (20 mM Tris-HCl pH8, 600 mM NaCl, 5% glycerol) and B (same as A plus 200 mM imidazole). Protein was bound to column at 10 mM Imidazole, washed at 25 mM, and eluted at 200 mM imidazole). The reasonably clean protein was dialysed against Storage buffer (20 mM Tris-HCl pH 8, 50 % glycerol, 200 mM KCl, 0.5 mM EDTA, 1mM DTT). DnaB was overexpressed from pCN24A in presence of Chloramphenicol, but likewise induced with 1 mM IPTG administered at OD₆₀₀= 0.8 and grown over night. The rest was done as with DnaG. NudC protein was purified similarly except in the absence of EDTA and DTT.

In vitro primer synthesis

Primers were synthesised by DnaG (+/- DnaB) on one of above mentioned templates in presence of 100 uM UTP and 10 uM GTP as well as approx 0.04 MBq α P³²-GTP per uL in primase buffer (50 mM HEPES pH 7.0, 20 mM Mg-Acetate, 100 mM K-Glutamate, 10 mM DTT) at 30° C for 10 minutes, unless otherwise indicated in results. The reaction was stopped either in Stop buffer (1x TBE, 7 M Urea, 20 mM EDTA, 100 ug/ml heparin, 0.02% Bromphenol blue, 0.02 % Xylene cyanole, 85% Formamide) or by addition of 1 M NaCl (final concentration). Buffer exchange or removal of excess NTPs and abortive products were performed by gel filtration (BioRad Micro-Spin 6), if necessary.

Pol I primer degradation

Primer was generated on the hairpin template (TTTACGCTTCGTTGACACACACACTGCGCGTTTGGGAAAACCTTTCCCAAA C) by DnaG (1 μ M) and DnaB (3 μ M) premixed at RT with initiator (100 μ M ATP or 1 mM other initiator (NAD, NADH, FAD, dpCoA)), 10 μ M (α -³²P) GTP (0.1 mCi/mmol), 100 μ M UTP, in primase buffer (50 mM HEPES pH 7.5, 20 mM Mg-Acetate, 100 mM K-Glutamate, 10 mM DTT) at 30 °C for 10 min. Reaction was stopped by adding a final concentration of 1 M NaCl. Protein was removed from reaction by adding Ni-NTA agarose beads in primase buffer and 1 M NaCl. After incubation at RT for 5 min, bead supernatant was transferred to gel filtration column (Micro-Bio Spin 6, BioRad). Filtrate was added onto DNA Polymerase 1 (PolI) (*E. coli* source, Fisher Scientific), 0.25U/uL reaction, alongside 10 uM dNTPs, and PolI buffer (20 mM Tris-HCl pH 8, 50 mM KCl, 10 mM MgCl₂). Incubation at 37°C was stopped after 5, 10, 20, 30, 45, 60, 120 and 180 seconds with Stop buffer.

References

1. Cahova H, Winz ML, Hofer K, Nubel G, & Jaschke A (2015) NAD captureSeq indicates NAD as a bacterial cap for a subset of regulatory RNAs. *Nature* 519(7543):374-377.
2. Jin Wang BLAC, Yong Lai, Hongping Dong, Luang Xu, Seetharamsingh Balamkundu, Weiling Maggie Cai, Liang Cui, Chuan Fa Liu, Xin-Yuan Fu, Zhenguo Lin, Pei-Yong Shi, Timothy K. Lu, Dahai Luo, Samie R. Jaffrey and Peter C. Dedon (2019) Quantifying the RNA cap epitranscriptome reveals novel caps in cellular and viral RNA. *bioRxiv* <http://dx.doi.org/10.1101/683045>.
3. Bird JG, *et al.* (2016) The mechanism of RNA 5' capping with NAD⁺, NADH and desphospho-CoA. *Nature* 535(7612):444-447.
4. Julius C & Yuzenkova Y (2017) Bacterial RNA polymerase caps RNA with various cofactors and cell wall precursors. *Nucleic Acids Res* 45(14):8282-8290.
5. Hofer K, *et al.* (2016) Structure and function of the bacterial decapping enzyme NudC. *Nat Chem Biol* 12(9):730-734.
6. Julius C, Riaz-Bradley A, & Yuzenkova Y (2018) RNA capping by mitochondrial and multi-subunit RNA polymerases. *Transcription*:1-13.
7. Swart JR & Griep MA (1995) Primer synthesis kinetics by Escherichia coli primase on single-stranded DNA templates. *Biochemistry* 34(49):16097-16106.
8. Zechner EL, Wu CA, & Marians KJ (1992) Coordinated leading- and lagging-strand synthesis at the Escherichia coli DNA replication fork. II. Frequency of primer synthesis and efficiency of primer utilization control Okazaki fragment size. *The Journal of biological chemistry* 267(6):4045-4053.
9. Wu CA, Zechner EL, Reems JA, McHenry CS, & Marians KJ (1992) Coordinated leading- and lagging-strand synthesis at the Escherichia coli DNA replication fork. V. Primase action regulates the cycle of Okazaki fragment synthesis. *The Journal of biological chemistry* 267(6):4074-4083.
10. Stayton MM & Kornberg A (1983) Complexes of Escherichia coli primase with the replication origin of G4 phage DNA. *J Biol Chem* 258(21):13205-13212.
11. Makowska-Grzyska M & Kaguni JM (2010) Primase directs the release of DnaC from DnaB. *Mol Cell* 37(1):90-101.

12. Sheaff RJ & Kuchta RD (1993) Mechanism of calf thymus DNA primase: slow initiation, rapid polymerization, and intelligent termination. *Biochemistry* 32(12):3027-3037.
13. Rymer RU, *et al.* (2012) Binding mechanism of metalNTP substrates and stringent-response alarmones to bacterial DnaG-type primases. *Structure* 20(9):1478-1489.
14. Bennett BD, *et al.* (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat Chem Biol* 5(8):593-599.
15. Zhou Y, *et al.* (2011) Determining the extremes of the cellular NAD(H) level by using an *Escherichia coli* NAD(+)-auxotrophic mutant. *Appl Environ Microbiol* 77(17):6133-6140.
16. Frick DN & Richardson CC (2001) DNA primases. *Annu Rev Biochem* 70:39-80.
17. Varik V, Oliveira SRA, Hauryliuk V, & Tenson T (2017) HPLC-based quantification of bacterial housekeeping nucleotides and alarmone messengers ppGpp and pppGpp. *Sci Rep* 7(1):11022.
18. Zerbe LK & Kuchta RD (2002) The p58 subunit of human DNA primase is important for primer initiation, elongation, and counting. *Biochemistry* 41(15):4891-4900.